CrossMark

ORIGINAL PAPER

How to get from the periphery into the core? The role of geographical location and scientific performance in network position in the field of neuroscience

Tamás Sebestyén[1] · **Orsolya Hau-Horváth**[2] ·
Attila Varga[1]

Abstract The main research question tackled in this paper is the role geographical position plays in scientific collaboration. In order to answer this question we ensemble a database containing the life-long publication records of best performing scholars in the field of neuroscience. Using this data we record the geographical location and the network position of these authors with respect to belonging to the center or to the periphery in both dimensions and use this information to analyze how moving into the core of the collaboration network is affected by being in the geographical core or periphery. A striking result from our analysis is that having double affiliations in the geographical center and periphery one can significantly improve performance in all respects (publications, citations, network position and knowledge availability), to even higher levels compared to authors exclusively locating in the geographical core.

1 Introduction

Innovation is an important factor in economic development and much of the innovation is based on scientific research—thus the evolution and continuous development of the

✉ Attila Varga
vargaa@ktk.pte.hu

1 MTA-PTE Innovation and Economic Growth Research Group and Department of Economics and Econometrics, Faculty of Business and Economics, University of Pécs, Pecs, Hungary

2 MTA-PTE Innovation and Economic Growth Research Group, Faculty of Business and Economics, University of Pécs, Pecs, Hungary

🖉 Springer

scientific research base is important not only from the scientific point of view but also for policy making. Innovation itself became a multi-actor process and the literature describes interconnected but nationally and regionally specific innovation systems (Nelson 1992; Braczyk et al. 1998). In this view innovation has become a collective process involving different actors like competing and related firms, supporting business services, private or public research organizations (Lundvall 2010). The increasingly interpersonal and inter-institutional nature of innovation is evidenced by the increasing role of co-authorship in the field of science.

Not surprisingly, there is a relatively wide literature focusing on the different dimensions of these co-authorship networks. Some studies focus on the topology of these networks, generally finding robust evidence on the scale-free characteristics of them, i.e. having large 'hubs' in the system with many co-authors while the majority of the authors have only a few co-authorship ties (Liu et al. 2005; Barabasi et al. 2002; Newman 2004). Other pieces of this literature provide several insights into the background of scientific collaboration, but there are some focal issues these studies concentrate on. One issue is the relationship between collaboration and performance which latter is typically measured by citation counts. There is evidence on the positive effect of degree centrality (Acedo et al. 2006; Abbasi et al. 2011; Yan and Ding 2009; Abbasi et al. 2012), closeness centrality (Yan and Ding 2009), betweenness centrality (Yan and Ding 2009; Abbasi et al. 2012), PageRank centrality (Yan and Ding 2009), as well as tie strengths (Abbasi et al. 2011) and access to structural holes as measured by Burt's network efficiency measure (Abbasi et al. 2012; Burt 1992). Interestingly Abbasi et al. (2011) find a negative effect for eigenvector centrality while Ahn et al. (2014) show that links to high-reputation universities also contribute positively to performance. Abbasi and Jaafari (2013) investigate the effect of the diversity of authors on publications' impact. They find that the geographical diversity of authors plays an important role in their research impact, but national collaboration better correlates with impact than international collaboration.

Another question, following directly from the previous one is if collaboration patterns determine performance, what determines collaboration patterns? A wide range of studies emphasizes the role of network distance (Fafchamps et al. 2010), technological proximity (Cunningham and Werker 2012), institutional proximity (Hardeman et al. 2012), academic excellence and informal communication (Jeong et al. 2011) and also similar affiliational background (Rodriguez and Pepe 2008).

There is another vein of literature, which primarily focuses on the geographical aspects of scientific collaboration (Varga 2009). Several studies support the crucial role of geographical proximity in innovation collaboration (Broekel and Boschma 2012; Cunningham and Werker 2012; Hardeman et al. 2012; Autant-Bernard et al. 2007). On the other hand, some authors argue that trust, frequent communication due to advanced ITC's or technological proximity can substitute for missing geographical proximity, and thus it is possible to establish collaboration even between distant participants (Boschma 2005; Hansen 2014). There is an increasing interest in the geographic patterns of scientific collaboration in itself: Maggioni et al. (2006), Hazir and Autant-Bernard (2013) or Varga et al. (2014) or Varga and Sebestyén (2017) provide insights into the relationship between interregional research collaboration and regional innovative performance.

However, the unit of analysis in geographical studies is the region and not the individual researcher. Once we analyze collaboration patterns at the individual level, the role of geography has a different dimension: actors may move between locations and as a result they develop (or deteriorate) collaboration networks as well as performance. Some recent studies focus on this issue by looking at the relocation of scientists from developing to developed countries. Saxenian (2006) and Agrawal et al. (2006) show that researchers emigrating from developing countries (China, India) to the USA positively affect their home countries' scientific performance through their intensive relationship with non-emigrating researchers. These results suggest that brain drain and brain-circulation are both present in science and the latter may significantly compensate for the negative effects of immigration.

In this paper we aim to contribute to the above-discussed veins of the literature. Our research framework is the interrelationship between geographical location, network position and the performance of individual scientists. Based on the discussion above, these three factors are tightly interrelated. We focus on the role of geographical location in the development of scientists' position in co-authorship networks. In order to handle this question we ensemble a database containing the life-long publication records of best performing scholars in the field of neuroscience. Using this data we identify the geographical and network positions of each author. We define core and periphery positions in both dimensions. Our target is to analyze the extent to which scientists' progress towards the core of the collaboration network is affected by their core or peripheral geographic location. We bring two interrelated questions into the picture. First, how geographical distance affects the results of scientific work: is physical proximity to the center of the scientific network necessary to become one of the best performing scholars? Second, what are the potential strategies of peripheral countries to catch up with core countries in scientific performance? Some exploratory analysis is carried out to shed light on the interrelationship between the three factors discussed here and then the focus is turned to the effect of geographical location on the evolution of network position.

The paper is structured as follows: in Sect. 2 we describe the database and introduce the measurement tools we employ. In Sect. 3 results of an exploratory data analysis on the relationship between performance, network position and geographical location are presented. In Sect. 4 we conduct an econometric analysis on the role of geographical location and performance in the development of scientists' network position. Section 5 summarizes the findings of the paper.

2 Data and measurement

This research builds on the database of Scopus, containing information on a wide range of scientific periodicals. We examined the possible use of different publication databases but finally decided to use the Scopus database as this is the largest available dataset on publications and citations indexing as many as 21 thousand different products of 5000 international publishers, including 20 thousand refereed journals.[1] In

---

[1] See: http://www.elsevier.com/__data/assets/pdf_file/0007/148714/scopus_facts_and_figures.pdf.

the following sections we describe how the final dataset under examination was built from this huge amount of data and then introduce the measurement of the concepts mentioned in the introduction using the data available in the database.

2.1 Data preparation

As the present research focuses on the field of neuroscience research, we restricted our search in the database for those journals which belong to the field of neuroscience. Scopus categorizes the different journals with respect to the scientific field they belong to and one source may be categorized under several different fields. We retrieved all journals from the database which were categorized under neuroscience, meaning 498 journals in total.

The main focus of this research is on the best performing authors in neuroscience, so we should further restrict the database with respect to authors in order to be left with really the top performing scientists in the field. This restriction was done in two steps: first, we restricted the journals under consideration, focusing on 'top' journals. Second, we restricted the authors for 'top' authors with the best publication performance.

In order to narrow down the list of journals, we took the SJR (SCImago Journal Rank) indicator as a basis.[2] The data we retrieved from Scopus contained information on 2008, 2009 and 2010 SJR scores, and we used the average of these 3 years for journal selection. We excluded those journals which did not have SJR score in the dataset or the score was zero for these 3 years:[3] this way 376 journals remained out of the 498 in the field of neuroscience. Then we ranked the journals according to their average SJR scores starting from the top journals and calculated the relative cumulative SJR score for each rank: e.g. a relative cumulative score of 15% at rank 5 means that the first 5 journals account for 15% of the sum of SJR scores for all journals. We decided to restrict our sample to those journals which account for 50% of the SJR scores which meant that we retrieved the first 57 journals in the field of neuroscience.

After the restriction of the sample to these journals, we retrieved the following information for all documents (publications) in these journals from 1974 to 2014,[4] available from the Scopus website:

- Last name of the authors and the first letter(s) of their first name(s)
- The title of the document (article)
- The title of the source (journal) in which the document is published
- The country of the affiliation of the authors (more than one affiliation is possible)
- The affiliation of the authors (more than one affiliation is possible)

---

[2] The SJR indicator is used for journal ranking which calculates the weight of the different journals according to the weight of those journals which cite that journal. See González-Pereira et al. (2010) for details.

[3] Our dataset contains SJR scores up to three decimals which may leave low-score journals with an SJR score of zero.

[4] Although the time of data retrieval was December 2013, the database contains information on some 2014 publications as well which were supposedly dated forward in the respective journals. However, we excluded 2013 and 2014 from the analysis as the information for these years must not be complete.

- The year when the document was published
- The number of citations received by the documents up to the date of the data retrieval (December 2013).

This restriction of the journals to the top 57 periodicals still contain around 340 thousand documents (publications) and altogether 1.5 million publication-author pairs. The time span of the database is quite large: we have publications from 1974 to the date of the data retrieval which is January 2013. This time window spans the entire life cycle of the scientific field under question as the first seminal publications in neuroscience date back around the time of our first year.

A large task after the data retrieval was to clearly identify authors to follow their publications. We have to be aware of misspelled names as well as the same author publishing under different names and also authors with the same name. We employed different methods to come up with an acceptable identification of authors. Due to the huge amount of data we employed algorithmic methods which are relatively fast but may leave some inconsistencies in the data. First, a character-distance method was used to identify possibly identical authors with similar names. Then, we used the affiliation data for the authors to search for further similarities in different author names and differences under the same author names. Finally we attached a unique identifier for each author—the dataset contains 370 thousand author identifier altogether.

If we are to focus on top authors, it is clear that having 370 thousand in the sample is simply too much, so we employed a second round of restriction narrowing down the sample of authors to the 'top' authors who publish the most according to the available data. This restriction is intuitively driven by the observation that out of the 370 thousand authors more than 50% has only one publication under the whole time span of the dataset, and more than 75% has at most 3 publications. Finally we restricted the sample of authors on those scientists who have more than 40 publications in total. This threshold was obtained by focusing on the top 1% of the authors. The logic of the restriction was similar to that used for the journal selection. The authors were ranked according to the number of their publications, starting from the highest publication record. Then we calculated the cumulative publication share for each author/rank. A cumulative publication share of 0.1% at rank 372 e.g. means that the first 372 authors account for 0.1% of the total publications in the dataset. Using this method we put the threshold where the authors in the list account for 1% of the total publications—this corresponds to 41 or more publications and the first most publishing 3838 authors, which is cc. 1% of the authors after identification in the 57 journals.

Finally, we have a dataset for analysis which contains the top scientists in neuroscience (with respect to their publication records), their publications in top journals in the field (with respect to SJR scores). This means 3838 unique authors and 57 journals.

2.2 Measurement and the final database

Using the available data outlined in the previous section, we try to analyze the interrelationship between geographical location, network position and scientific performance. In order to quantify these concepts, we apply the following measurement techniques using the data available from the database we constructed.

2.2.1 Geographical location of an author

The database contains two sources of information on the geographical location of the authors. First, we have a country code for each author and for each publication—one author may have several countries at the same time if he or she has multiple affiliations.[5] Second, we have data on the exact affiliation which would allow us in principle to provide a detailed (sub-country) identification of the authors' location. However, the affiliation data is available in character format, and it would need vast clearing to bring this information into a usable format. On the other hand, even using countries as the basis for geographical localization would overwhelm the analysis, we decided to group the countries into two groups: the geographical center and the geographical periphery.

However, the categorization of countries into the geographical center or the geographical core is not straightforward. We use an approach which is based on the occurrence of countries as affiliation for scientists in the field of neuroscience by saying that those countries belong to the geographical center where most of the publications are created. As setting the threshold in this respect may be crucial, we use four different categorizations from broader to narrower definitions of the geographical center, as listed below.[6]

- *Broad definition*. In this case we include US, EU15 countries, Japan, Canada, Australia, Switzerland, Norway and Israel as geographically core countries. On average these countries account for more than 90% of the author affiliations in our sample.
- *Europe definition*. In this case we include EU15 countries plus US and Japan as geographically core countries. These countries account for more than 80% of the author affiliations on average.
- *Narrow definition*. In this case we include US, UK, Japan and Germany as geographically core countries. These four countries account for more than 60% of author affiliations on average.
- *Strict definition*. In this case only the US is counted as the geographical core. US is always the top country with respect to publications and account for more than 40% of the author affiliations on average.

Using this classification we establish two measurements for the geographical location as described below:

- *Discrete location*. According to this measurement an author is characterized into three possible categories: (i) having affiliation only in countries classified as the geographical center; (ii) having affiliation only in countries classified as the geographical periphery; (iii) having affiliation in both regions (this is labelled as 'double affiliation' later in the paper).

---

[5] In some cases the country of the affiliation of authors was missing from the retrieved data. For many such cases, though, manual clearing was possible to find the respective country codes. We did not find a country of the affiliation for only 2% of the records, which were treated as missing data in the final database with respect to location, but other characteristics of the given author may be available for the same record.

[6] Our results are quite robust with respect to the choice of the definition of geographical center.

- *Continuous location*. According to this measurement an author is characterized by a value between 0 and 1 which measures the percentage of his or her affiliations which belongs to the geographical center. The higher this number, the more exclusively an author is working in the geographical center.

With respect to our main research question regarding the role of geographical position in network position and scientific performance, these measures of geographical location are able to reflect connections in physical proximity: if an author is categorized as double affiliated with respect to the discrete measure or a value between 0 and 1 with respect to the continuous measure, then it means some kind of physical proximity between the geographical core and periphery, while exclusive affiliation in either the geographical center or core means physical distance although the core and periphery may be linked together by co-authorship (a-spatial, network) links even in this case.[7]

2.2.2 Network position of an author

Using the information retrieved from the Scopus database we are able to build a co-authorship network of neuroscientists for all years of the sample. The network we use is a weighted network where the weight of a co-authorship tie is based on the number of publications the two authors had together. As we have a long time span, the co-authorship network is constructed for each year separately. In formal terms let $a_{i,j,t}$ denote the weight of co-authorship link between authors i and j in year t. This weight is simply the number of publications which authors i and j co-published in year t. Having this raw view on annual publication records can be misleading with respect to the duration of interpersonal links: links are only recorded for the year when the publication is published. It is intuitive to think, though, that several years of work precedes this publication, so the interpersonal relationships exist even before the publication date. In order to account for this bias, we use a 5-years moving aggregation. This means that in the final network data, used for the analysis later on, the link weight between two authors is the sum of the raw link weights in the given year and the four preceding year. Formally, we define $b_{i,j,t} = \sum_{\tau=0}^{4} a_{i,j,t-\tau}$ and $b_{i,j,t}$ is used in the analysis—so the link weight of a given year is the sum of co-publications in that year and the four consecutive years. This way we have cooperation data for every author and for every year, but this data reflects a smooth evolution due to this aggregation.

Using this network data we can calculate measures which capture the authors' position in the co-authorship network.

We operationalize network position by using two different measures as described below.

- *Coreness profile*. Following the method described in Della Rossa et al. (2013) we assign a value ranging from 0 to 1 to each author which reflects the extent to which the given author belongs to the core of the network. The higher this value, the more connected the author is and the lower, the more peripheral he or she is.

---

[7] Note that physical proximity between geographical areas is not a fixed concept here: physical proximity is driven by authors relocating and maintaining affiliation in both areas, so these locations become physically proximate through the authors moving between them.

- *ENQ index*. Following the method of Sebestyén and Varga (2013a, b) described in the Appendix we assign a value to each author, which measures the quality of available knowledge from their co-authorship networks.

The difference between the coreness profile and the ENQ index is that while the former takes into account the connection structure of the network, the ENQ index also accounts for the knowledge level of the direct and indirect co-authors. On the other hand, while the coreness profile is normalized to the interval between 0 and 1, the ENQ index is unbounded and driven by the total knowledge in the network.[8]

In addition to measuring pure network position we also constructed two measures in order to capture the structure of links an author has. When looking at the evolution of networks, it is not only important to know what the position of the author is in the network but also to whom they are linked. We can introduce a simple measurement of this in the context of our categorization of geographical center and periphery versus network core and periphery. In these measures network position is characterized by the characteristics of the partners.

- *Link share to the network core*. In this measure we simply calculate the share of links of an author which point to co-authors who are in the core of the network, relative to all links of the author. This allows us to measure the extent to which an author's links focus on core scientists.
- *Link share to the geographical center*. Analogously to the previous measure, here we calculate the share of links of an author which point to co-authors who are in the geographical center, relative to all links of the author.

As mentioned in the case of proxies for geographical location, those measures can reflect physical proximity between the geographical center and periphery, transmitted by relocating scientists. In contrast, the network measures reflect links which can be maintained in physical distance (but network proximity does not necessarily coincide with geographical distance).

2.2.3 Performance of an author

The third concept we intend to measure in this study is the scientific performance of the authors. We give three possible measures for the performance of an author, based on the information available from the database:

- *Publications*. We simply count the number of publications of an author in a given year. This is the most common performance measure.
- *Citations*. We simply record the number of citations obtained by a given author. Although this is also a widely used measure of scientific performance, the limitations of our database call for some attention when using these records. The available information is the citation count of a paper up to the point of data retrieval. We can then provide an annual measure for all authors: the citation value of an author

---

[8] As it is described in the Appendix together with Sebestyén and Varga (2013a, b), the ENQ index requires the operationalization of the knowledge levels of the authors. In this paper we use the cumulated number of publication as a proxy for individual knowledge levels.

i in year t is the number of citations on the papers published by author i in year t (up to the date of data retrieval). However, these counts must be biased at least for two reasons: (i) earlier publications have more time to accumulate citations, so citation counts of younger authors with most of their papers in recent years are biased downwards; (ii) citation counts even for one single author are not necessarily comparable for the same reason: older publications may accumulate more citations.

- *Citations over publications*. We simply divide the number of citations an author received with the number of paper he or she published in a given year. This can be regarded as an efficiency measure: what is the impact an author can reach through one publication. However, the problems of the citation counts are also valid in the case of this measurement.

Although we are aware that these measures of scientific performance are far from perfect and there are better candidates (such as the H index among others), our dataset contains information only on these indicators and retrieving further performance measures and compiling the with our dataset would be a serious work, left for further research.

2.2.4 The final database

Once these measures are calculated, we assembled the final database for analysis. As we have many observational units (authors) for many years, we can render the data into a panel database. However, in order to make comparisons between authors, it seems to be more interesting to render the panel database on a career-year basis rather than on a calendar-year basis. This means that we take each author, record the first year of his or her activity in the database (proxied by the first publication) and then this year is going to be his or her first year in the database. This way the database contains in the first period the first career-year of all authors and so on. Although the personal comparison is obtained here, the panel becomes unbalanced as there are a few authors whose life cycle span almost the whole period (around 40 years) but many of them have shorter careers. Another source of bias is that although some careers are full in the sense that we have early and late career records for these individuals, but many careers are truncated as these authors are still in their early or middle career years at the date of data retrieval.

3 Exploratory analysis

In this section we provide and discuss some descriptive statistics of the data established in the previous sections. The general logic behind these statistics is that we split the sample of authors according to some characteristics (e.g. locating in the geographical center or in the geographical periphery) and then see if the different subsamples have different patterns with respect to some variables (e.g. performance, network position) over the career years. First we show some analysis with respect to the role of geographical location (i.e. when subsamples are made according to the geographical location)

Table 1 Summary statistics of the data over different subsamples

	Location			Sum
	Periphery	Double	Center	
Network				
Periphery				
Nr. of observations	12,227	2755	32,155	47,137
Average publication	1.98	3.96	2.01	2.12
Average citation	81.22	158.84	100.39	98.84
Average ENQ	15,765	98,024	23,800	137,588
Core				
Nr. of observations	9020	5460	27,317	41,797
Average publication	3.44	7.81	3.73	4.20
Average citation	158.27	275.80	208.22	206.27
Average ENQ	21,245	152,485	42,758	216,488
Sum				
Nr. of observations	21,247	8215	59,472	88,934
Average publication	2.60	6.52	2.80	3.10
Average citation	113.93	236.57	149.92	149.33
Average ENQ	37,010	250,508	66,558	35,4076

and then on the role of network position (i.e. when subsamples are made according to network position).

Table 1 gives a brief summary of the data we use. In this table we sort the full sample with respect to two dimensions, also used in the following analysis: geographical location and network position. With respect to location, we differentiate according to the discrete measure as described above, and with respect to network position, according to the coreness profile measure (a coreness score above 0.5 is counted as core, and below 0.5 is counted as periphery). The table contains the number of observations in each possible category, the average publication, average citation and average ENQ index for these observations.[9] In the table we use the Narrow definition of geographical center, but the numbers are qualitatively similar for the other definitions.

3.1 The role of geographical position

In our very first attempt the sample of authors are divided into two groups according to their life-time geographical position. We took the *continuous location* measures—each author has one value for all career years—and calculated the individual averages over each author's lifetime. These values reflect that to what extent an author belongs to the

---

[9] The unit of observation is a given author in a given year due to the panel structure of the data. Coreness scores and location measures are not recorded as these values serve as the basis for the categorization of the observations.
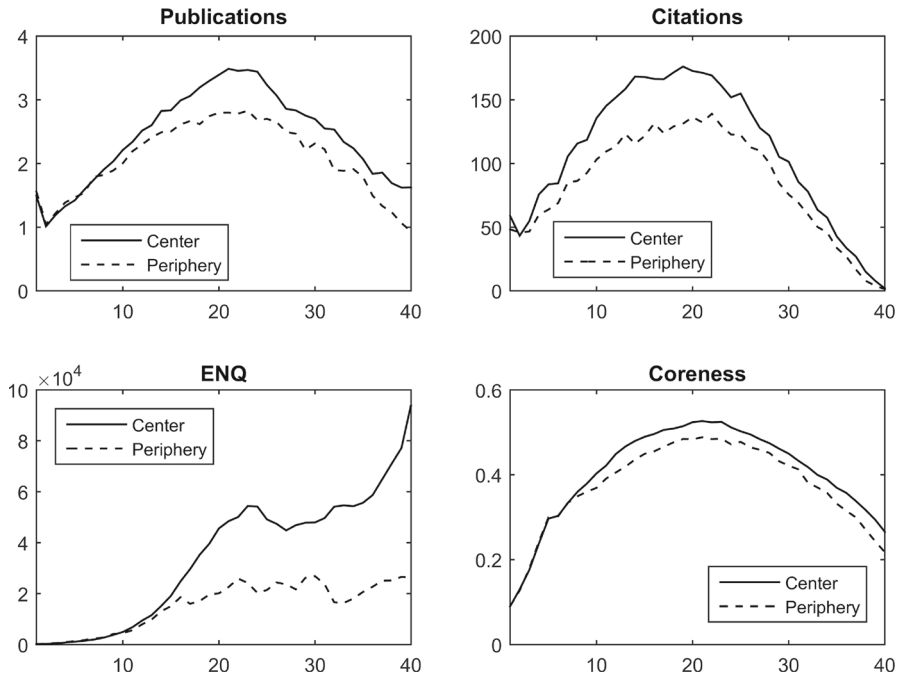
Fig. 1 The role of main lifetime geographical location on performance and network position

geographical center or periphery over his or her lifetime.[10] We then split the sample between 'mainly center' and 'mainly periphery' authors, the former group containing those authors whose average location value is above (or equal to) 0.5 and the latter group containing those whose value is below 0.5. As defined previously, we apply four different categorizations for the geographical position of an author (country). In the diagrams and consequent analyses below we refer to the Narrow definition, but all other categorizations yield qualitatively similar results, so as mentioned earlier, the way we define the geographical core (within reasonable limits) does not affect the patterns we find in the data.

Figure 1 collects the data with respect to this analysis. The horizontal axis on each panel corresponds to career-years. The solid lines represent the average publication, citation counts, ENQ and coreness values for those authors who on average belong to the geographical center over their lifetime and the dashed lines represent the average values of those authors who belong to the geographical periphery over their lifetime. In each panel the title defines the value which is measured on the vertical axis.

---

[10] In contrast to 'current geographical position', which reflects the affiliation an author has in a given year, the 'lifetime' or 'average' position is the dominant position she reports as affiliation through her entire lifetime. For example, if an author is affiliated in the US for 5 consecutive years, then moves to Mexico for 2 years and then back to the US for another 3 years, her current position is 1 for 5 years (being in the geographical center), 0 for the next 2 years (being in the geographical periphery) and then again 1 for the following 3 years. Assuming that this is her entire lifespan in the sample, her lifetime position is the average of these, i.e. 0.8. Which means that she is dominantly affiliated in the geographical center.

Some observations are straightforward here. Except for the ENQ index, all measures exhibit a reversed-U shape which nicely shows that the researchers in the sample are the most active (have more publications and better network positions) in their mid-career while during their early and late career they are less active in publications, publications in these life stages receive less citations and they move out from the core of the network. On the other hand, the ENQ index shows a different path: increasing throughout the time window which is due to the fact that this index is driven by the knowledge level measured by cumulated publications which increase over time by definition. On the other hand, observing the contrast between the coreness profile and the ENQ index we can argue that although the authors move out from the network core in their late careers they manage to maintain important connections which link them to the most knowledgeable partners.

With respect to the difference between the two subsamples (authors mainly in the geographical center and geographical periphery) we can also see some important results. First, it seems that the authors' overall or typical geographical location significantly affects performance: those authors who belong mainly to the geographical center have more publications (excluding the initial years of the careers), much better citation performance in their mid-career. Second, geographical location also differentiates with respect to network position. If we take the coreness profile, we see that authors in the geographical center are better positioned in the network in their mid- and late career, but in the early careers the difference is not significant (similarly to publications). Third, if we look at the ENQ index, we see the same picture: in the mid-career phase those authors have typically higher ENQ index who mainly belong to the geographical center. Although the difference between center and periphery authors dampen (citations) or stabilize (publications and coreness) with respect to other measures in the late career years, this difference in the ENQ index becomes even larger which means that authors with dominantly peripheral affiliations typically lose their best connections to the network while authors dominantly in the geographical center typically increase their access to knowledge up to the end of their career.

In Fig. 2 we show a similar analysis, but the differentiation is now done on the basis of the current location of the authors. While in the previous analysis each author had one category according to which she belongs to the geographical center or periphery on average over the lifetime, now each author has a category for each year of her career depending on the affiliations she has in that given year. In contrast to the previous analysis where we used the continuous location measure, now we base the analysis on the discrete location measure. According to this all authors (in all years) are classified into three categories: affiliated only in the geographical center (dashed lines), affiliated only in the geographical periphery (dotted lines) or affiliated both in the center and the periphery (solid lines). The main features of the data are also reflected here as in Fig. 1: curves are inverse-U shaped except the ENQ index and authors affiliated in the geographical center outperform periphery authors in citations, coreness, ENQ and slightly in publication. What is striking that authors with double affiliation (both in the geographical center and geographical periphery) even outperform those only affiliated in the geographical center, and this is true for all four indicators and for all career years. Moreover, in the case of publication, ENQ and coreness, those with double affiliations do not seem to decrease publication activity and loose network
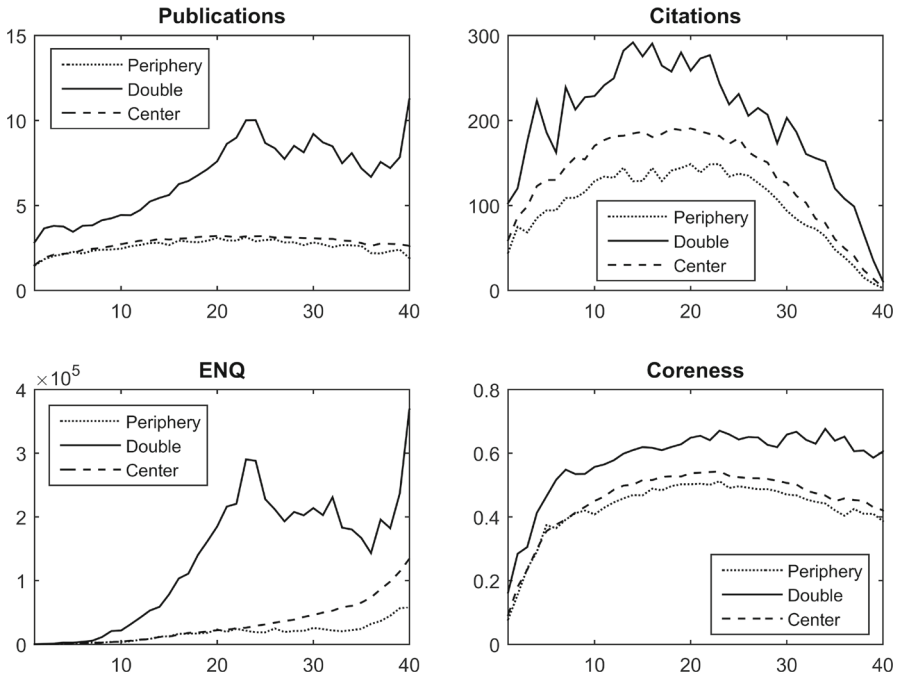
Fig. 2 The role of current geographical location on performance and network position

position even in the late years of their career as 'single-affiliated' authors seem to do. These results show that moving between the geographical center and periphery is able to positively influence network position and publication performance especially if the former affiliation is maintained.

It is also interesting to see whether the initial geographical location of the authors makes any difference in their performance and network position over their career. We checked this question by focusing on those authors who have moved between the geographical center and periphery over their career at least once and grouped them according to their initial location. The results with respect to publications can be seen in Fig. 3.[11] The graph shows no significant difference in the performance of authors who start their career in the geographical periphery and those in the center, not even in the early career years. However, we must call attention to the fact that these results only reflect the career of the sample authors within the filtered top journals, which means that we may have omitted information on the career of authors before they start to publish in the top journals.

---

[11] The initial location of each author is determined on the basis of our continuous location measure. We took the first 5 career years of each author, averaged their continuous location scores over these years and if this average was above 0.5 they were classified as starting in the geographical center and if it was 0.5 or below, they were classified as starting in the geographical periphery. As before, the results are shown for the Narrow definition of the geographical center, but the results do not change considerably if we use alternative definitions for the geographical core or the time window to define initial position varies (between 1 and 5 years).
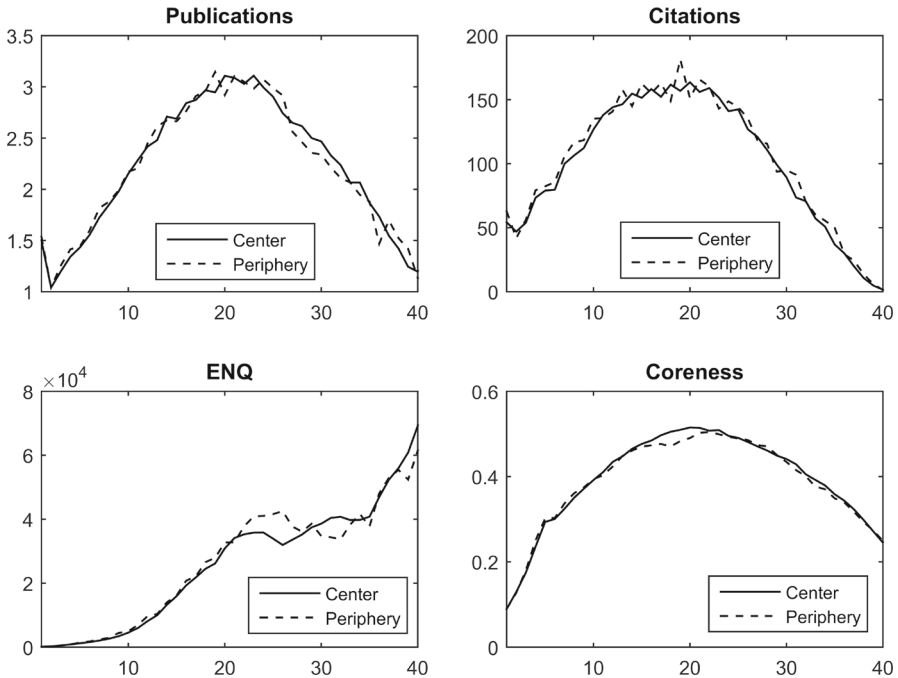
Fig. 3 The role of initial geographical location in performance and network position

3.2 The role of network position

In the previous section we looked at the role of the geographical location of authors
on their performance and network position. In order to further examine the interrela-
tionship between the three concepts as introduced in the introduction, now we look at
how being at different network positions affects performance and geographical loca-
tion. In other words, the sample of authors is now split between those authors who
are in the network core and those who are at the network periphery and we examine
whether there is any significant difference between the two groups with respect to their
geographical location and performance.

In Fig. 4 the sample is split along life-time network position. We took the coreness
profile of each author for every year and calculated the average coreness value for all
authors over their lifetime. Then all authors with a 0.5 or higher value were classified as
belonging to the network core and those with lower than or equal to 0.5 were classified
as belonging to the network periphery as their main or average network position.[12]

The upper two panels of the figure read analogously to the previous figures. The
dashed lines show the average publication and citation values of authors belonging
to the network periphery on average while the solid lines that of the authors in the
network core. These figures reflect again the inverse-U shape as before and show a

---

[12] Although the co-authorship network is scale-free with a power law degree distribution, the coreness
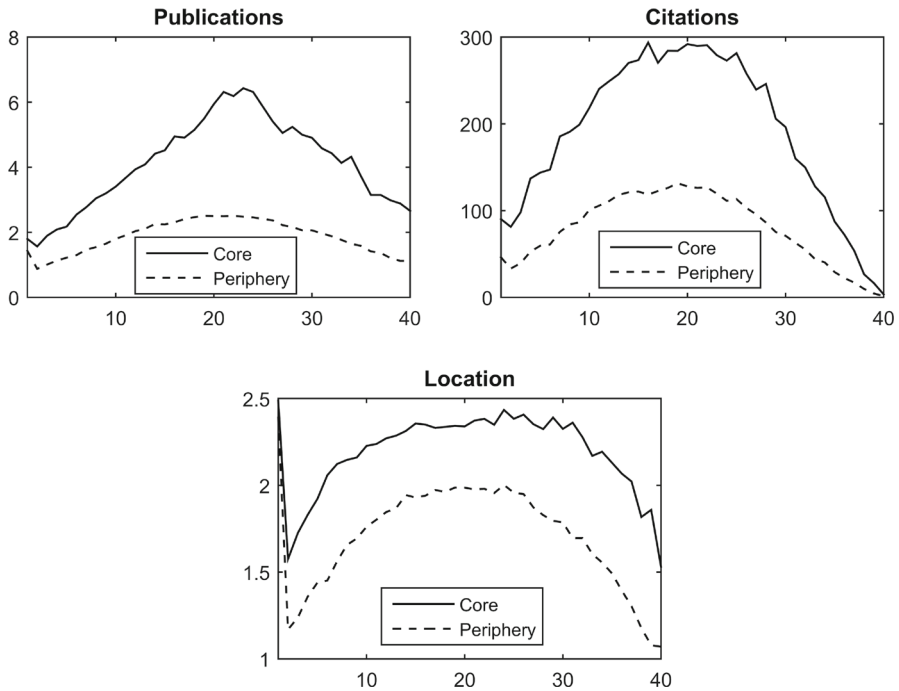values are scattered symmetrically around 0.5.

Fig. 4 The role of main network position on performance and geographical location

significant difference in performance in favor of the authors who maintain their core network position over their lifetime. It is interesting to see that the relative difference between the two groups of authors is around two-fold overall the authors' career and in both performance measures.

The bottom panel was constructed using the discrete location classification of authors. We assigned a value of 1 to those authors who are affiliated only in the geographical periphery in a given year, a value of 3 to those authors who are affiliated only in the geographical center and a value of 2 to those authors who have affiliation in both the geographical center and the periphery. As the two lines represent average values for the two sub-samples (network core and network periphery authors), the values on the vertical axis reflect the extent to which authors in the given subsample are located in the geographical center, periphery or in between on average. In this panel we observe, similarly to the other two panels, that authors who spend the majority of their career positioned in the core of the network are those who are located mainly in the geographical center.

Figure 5 reflects the differences in network position but instead of the average lifetime position, now the current, annual position is the basis for splitting the sample of authors. The results here reinforce those obtained from Fig. 4: a core position in the co-authorship network positively affects performance and better connected authors are more likely to dominantly be located in the geographical center. According to Fig. 5 this is not only true on average over the lifetime of the authors but also on an annual
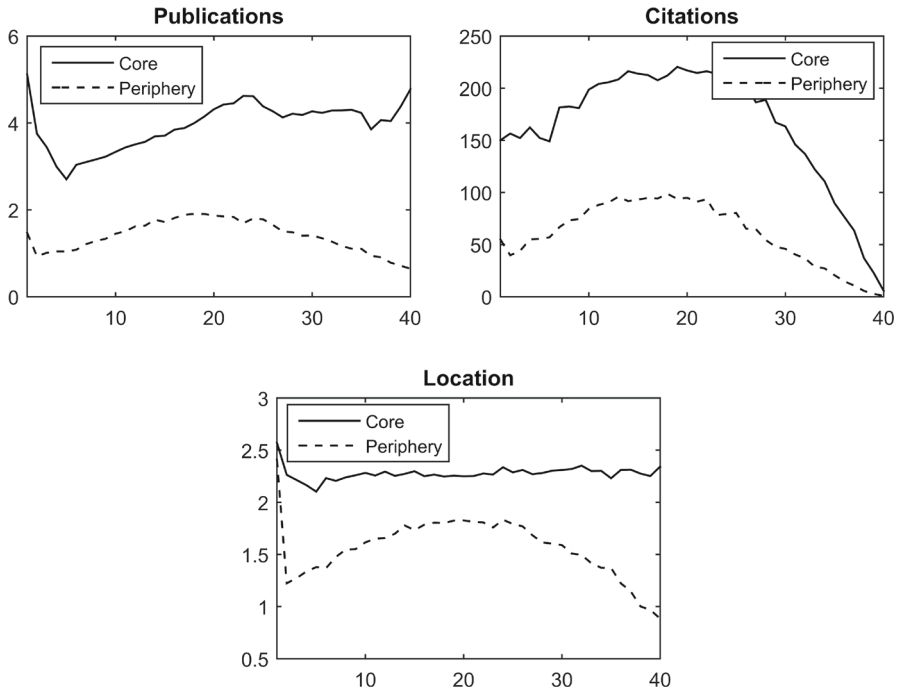
Publications

Citations

Location

Fig. 5 The role of current network position on performance and geographical location

basis. In addition to this, with respect to publications and location, in Fig. 5 we do not observe the inverse U-shape in the case of those authors who belong to the network core. This means that those authors who maintain their core network position in their late career years, typically remain in the geographical center and maintain publication activity comparable to their mid-career years.

3.3 Link formation processes

In the previous sections we focused on the position of the authors in the network and their geographical location as well as their performance. In what follows, our focus is not only on the position of the authors but where their links point and how this is related to their position in the network and their geographical location. In order to operationalize this, we calculated for each author and for each year the share of links which points to the geographical center and also the share of links which point to the network core.[13]

---

[13] Let $L_{it} = L_{it}^P + L_{it}^C$ be the number of coauthorship ties scientist i has in year t, which is split among ties where the coauthor is affiliated in the geographical center (L_{it}^C) and ties where the coauthor is affiliated in the geographical periphery (L_{it}^P). The link shares towards the geographical center and periphery are then respectively: L_{it}^C/L_{it} and L_{it}^P/L_{it}. We construct link shares to the network core and periphery by a similar logic.
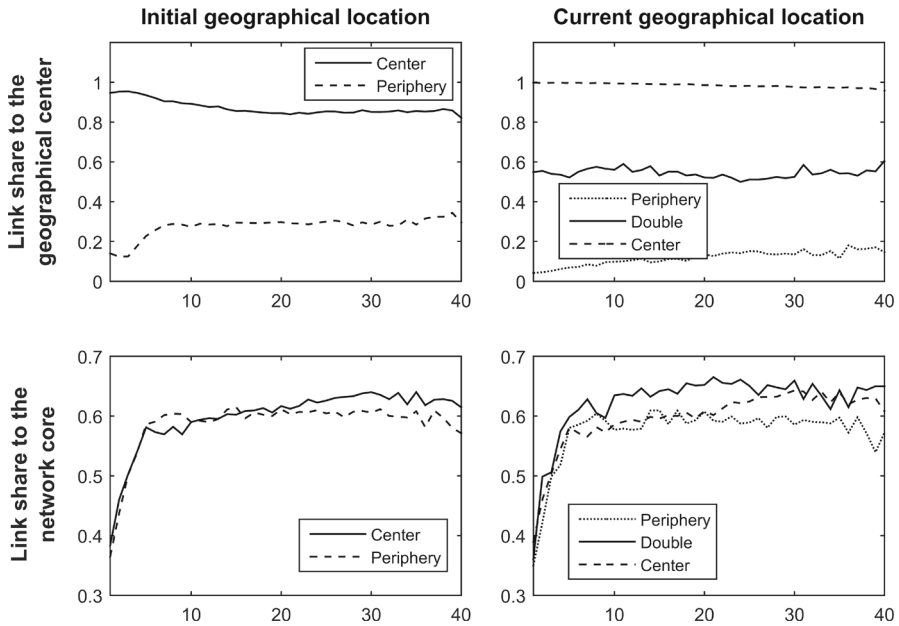
Fig. 6 The role of initial and current geographical location in the share of links pointing to the geographical center and the network core

Figure 6 provides some insight into the relationship between geographical position and the direction of links with respect to network and geographical positions. The figure reads similarly to the previous ones: each panel shows average link share values for two and three subsamples, differentiated along the initial and current location of authors.[14] The upper two panels have the link share pointing to the geographical center on the vertical axes while the bottom two panels have the link share pointing to the network core on the vertical axis. On the panels in the left column the sample of authors is split according to their initial geographical location whereas on the right panels the current annual location is the basis of the subsamples.

With respect to the link share to the network core the first observation is that there is no huge difference between the subsamples irrespective of the basis of the categorization. All authors start at the relative periphery of the network, around 30-40% of their links pointing into the network core and then relatively rapidly they develop their position with around 60% of their links pointing into the network core already at the end of the first decade of their career. Authors starting in the geographical center maintain a slightly higher share of their links towards the network center at the late career years, but this difference is not significant. If we look at the subsamples

---

[14] As before, initial location was determined on the basis of the continuous location measure, while current location was determined on the basis of the discrete location measure. As a result, for initial location we have two categories, and for current location three categories, including the double affiliation. The figure shows the results for the Narrow definition of geographical center, but the results are similar for the other definitions.

based on the current geographical location (bottom right panel) we see again the differentiating role of double affiliations. While in the case of peripheral authors less then 60% of their links point to the network core, this share is around 65%, a bit (but statistically significantly) higher for double affiliated authors, which may show part of the explanation behind their overperformance recorded previously. This observation is augmented by the fact that authors in the geographical center do not outperform those in the periphery up to the mid career years, but catch up with double affiliated authors at the late career years.

On the other hand, the picture is very different when we focus on the geographical arrangement of connections (upper panels). The first striking point is that in spite of the fact that we did not find any effect for the initial geographical location on performance and network position (see Sect. 3.1) and only a very weak effect on the share of links pointing into the network core (see the bottom-left panel of Fig. 6), there is a marked effect of initial location with respect to the share of links pointing to the geographical center. It is interesting to see that those authors who start from the geographical center have links which almost exclusively point to the geographical center. There is a slight change in this pattern over the career years but throughout their life time these authors are connected almost exclusively to the geographical center. On the other hand, authors starting at the geographical periphery have a small amount of links toward the geographical center (below 20%) but they increase this share continuously and at late career years this share stops slightly below 40%. This shows that the initial location largely determines the network connectedness of authors with respect to its geographical scope. Authors starting from the center heavily rely on coauthorship links within the center throughout their career, and the opposite is true for authors coming from the periphery.

A very similar picture is drawn on the upper-right panel where authors are grouped according to their current geographical position: this position in a given year affects the share of links towards the geographical core in the same year. There is a very clear difference between authors affiliated only in the geographical center who almost exclusively maintain links with other authors also affiliated in the geographical center, while the case is reversed for authors affiliated only in the periphery. It is nice to see that double affiliated authors have a quite balanced network position with around 60% of their links pointing into the geographical center. These differences between the three subsamples do not change too much with the age of the authors, although the share of center authors slightly decrease to the end of their career, while peripheral authors increase their share a bit more considerably.

3.4 Discussion

In the previous sections we presented some descriptive findings with respect to the co-evolution of geographical location, network position and scientific performance. First, the results indicate that within the selected top performing authors we can find scholars from the geographical center as well as from the geographical periphery. In addition, authors in the geographical center seem to have more publications and be more cited compared to their peers in the geographical periphery (Figs. 1, 2).

Scientists naturally start their careers at the periphery of the network regardless of their geographical location, but in their later career they also typically lose their connectedness (Figs. 1, 2), which is also reflected by the fact that authors in the network core typically belong more to the geographical center (Fig. 4). With respect to the knowledge authors can access through their network (the ENQ index), we also see geographically central authors to outperform their peripheral peers. On the other hand, geographical location does not seem to heavily affect the share of links, which point towards the network core, but both initial and current location has a massive impact on the geographical distribution of the links an author has.

We may find two important results from these facts. First, with respect to scientific embeddedness (network position) we can say that having good links (as reflected by the ENQ index and by the link share to the network core) is comparable to being in the geographical center with regards to publication: authors who are in the geographical periphery can have similar publication patterns as those in the center. On the other hand, these authors have typically less citations and their network embeddedness is less favorable—they are typically in the periphery of the network. Second, in order to improve performance and embeddedness (citations and network position as measured by being in the core of the network) one can not skip physical interactions. A striking result from the descriptive analysis is that by having double affiliations in the geographical center and periphery one can significantly improve performance in all respects (publications, citations, network position and knowledge availability), to even higher levels compared to authors exclusively locating in the geographical core. So it is not only important to improve scientific collaboration in order to get into the network core from the geographical periphery but also to maintain physical proximity by working in both geographical locations at the same time.

4 The determinants behind improving network position: a regression analysis

In the previous sections we heavily focused on some descriptive facts provided by our dataset about the interrelationship between network position, geographical location and scientific performance. In this section we use the panel structure of our dataset in order to carry out some regression analysis where we try to explain the evolution of network position with the other two concepts, namely geographical location and performance. One of the most important results of the previous section was that having double affiliation significantly contributes to performance and network position (Fig. 2). In this section we revisit these descriptive findings with the help of regression analysis.

As indicated by the title of this paper, our main focus is on scientists' career with respect to the evolution of their network position. As mentioned in the Introduction, there are several papers which try to explain the performance of the scholars with their network position and they find a generally significant relationship between the two. However, the direction of the relationship between network position and performance is not evident: it is highly possible that these characteristics co-evolve with one reinforcing the other. As a result, it is straightforward to ask whether network position

is affected by performance and geographical location. In the following analysis, we concentrate on the latter issue: we search for the determinants of improvements in scientists' network position and we try to trace the role for double affiliation in this story.

We employ a dynamic panel specification of the following general form:

$$NWPOS_{i,t} = \alpha + \delta \cdot NWPOS_{i,t-1} + \beta_1 \cdot GEOLOC_{i,t} + \beta_2 \cdot PERF_{i,t} + u_i + \varepsilon_{i,t}$$
(1)

where $NWPOS_{i,t}$ is a proxy for the network position of author i in career year t, $GEOLOC_{i,t}$ is the geographical location of author i in career year t while $PERF_{i,t}$ is a proxy for the performance of author i in career year t. u_i is a unit (author-) specific fixed effect while $\varepsilon_{i,t}$ is an observation-specific error term. Using this formulation we can examine whether geographical location and scientific performance affect the improvements in network position: the lagged network position on the right hand side controls for the persistence in network position over time, and also drives the regression to measure the effect of the other explanatory variables on the change in network position compared to the previous period. Equation (1) is a general form which is operationalized with different measurements. Network position can be measured by the coreness profile and also by the ENQ index as introduced in Sect. 2.2 Geographical location can be proxied by the discrete and continuous measures, both calculated along the four different definitions of the geographical center. Finally, performance is reflected by publication, citation counts and the ratio of these two measures. In the following analysis we present results only for the case where both publications and the efficiency measure of citations over publications are included. Further results, not presented here, show that these performance measures are orthogonal so including the separately or together into the regression does not yield different results.

Table 2 contains regression results when the coreness profile is used as the $NWPOS_{i,t}$ variable in Eq. (1). The model settings are summarized in italics. We run regressions using the continuous and the discrete location measures, both calculated according to the four definitions of geographical center. In case of using the discrete location measure, we employed a dummy variable approach to enter these discrete categories into the model. $DISCLOC\_double_{i,t}$ is a dummy variable which takes the value of 1 if author i in period t has double affiliation. $DISCLOC\_center_{i,t}$ takes the value of 1 if author i in period t is affiliated only in the center, while $DISCLOC\_none_{i,t}$ indicates if there is no information on the geographical position of author t in period i. Using this setting our reference group are those authors who are affiliated only in the geographical periphery. In case of using the continuous location measure we also include the squared version of the variable in order to see whether there is an inverse U-shaped relationship between location and network position indicating that belonging exclusively to the geographical center or periphery is less favorable compared to the double affiliation. We use publication counts and the citation over publication measure of efficiency as proxies for scientific performance. As there is no significant correlation between the two performance measures we include them together into the regression. Finally, we control for the age of authors by including the career year as

a further explanatory variable ($CARR\_Y_{i,t}$) as well as the squared form in order to account for the higher performance of mid-career authors. Table 2 contains of the regression results for these settings.

The first thing to note in the regression results is that the lagged values of network position are always significant and positive, indicating a string persistence in network position over time. Second, publications and the efficiency of publication is always positive and significant, so it is reinforced by these results that a better performance contributes to network position: more publications and more citations per publication lead to a better connected network in terms of the coreness profile. It is interesting to add that the citations per publication measure is an ex post variable in the sense that it reflects the efficiency of the authors' publications in a given year as measured at the time of data retrieval—this means that although the number of citations on a publication takes time to increase to high levels, this kind of 'quality' is reflected almost immediately after publication by an improved network position. Third, we find a negative significant coefficient for the squared career year variable which indeed indicates a hump-shaped relationship between network position and age as evidenced by Figs. 1 and 2. When geographical location is proxied by the discrete measure (dummy variables, Models 1–4), we see a positive significant coefficient for double affiliated authors, which means that this group of authors has better network position in terms of coreness, compared to peripheral authors. Authors who are located exclusively in the geographical center are either not outperforming peripheral authors (Models 1 and 4 with the Broad and Strict definition of geographical core) or they do (Models 2 and 3 with Europe and narrow definitions for geographical core) but in the latter case the magnitude of the coefficient for double affiliated authors is much higher which indicates that although authors who are affiliated in the geographical center may have better network position than those in the geographical periphery, double affiliated authors (with affiliation both in the geographical center and periphery) even outperform exclusively center authors. This result reinforces the visual impressions obtained in the descriptive analysis. The regressions including the continuous measure for geographical location (Models 5–6) also reflect the same thing: the negative significant squared term indicates an inversed U-shaped relationship between location and network position. According to this those authors can improve their network position who have affiliations both in the geographical center and periphery. Authors who are affiliated exclusively in the geographical center or periphery have less favorable network positions. When location is measured with the continuous variable, this finding is strongly significant and robust for all definitions of the geographical core.

Table 3 presents regression results when network position is measured by the ENQ index as presented in Sect. 2.2 and the Appendix. The results reinforce the impressions discussed previously. Network position measured by the ENQ index is also persistent, publication and efficiency have a positive impact on network position and there is a significant positive role for 'intermediate' geographical positions i.e. when an author is not exclusively affiliated in the geographical center or periphery—this latter effect is present irrespective of the proxy for geographical location and the choice of the geographical core. The only difference is that the coefficient of squared career years is now positive and significant which indicates a positive (and nonlinear) trend in ENQ over the authors' lifetime which is also evidenced by Figs. 1 and 2.

Table 2 Regression results for coreness profile

	Model(1)	Model(2)	Model(3)	Model(4)	Model(5)	Model(6)	Model(7)	Model(8)
NETPOS	Coreness	Coreness	Coreness	Coreness	Coreness	Coreness	Coreness	Coreness
GEOLOC	Discrete	Discrete	Discrete	Discrete	Continuous	Continuous	Continuous	Continuous
GEOCENT	Broad	Europe	Narrow	Strict	Broad	Europe	Narrow	Strict
PERF	Pub & C/P	Pub & C/P	Pub & C/P	Pub & C/P	Pub & C/P	Pub & C/P	Pub & C/P	Pub & C/P
NETPOS(−1)	0.6553***	0.6551***	0.6547***	0.6548***	0.6550***	0.6550***	0.6547***	0.6548***
	(0.0065)	(0.0065)	(0.0065)	(0.0065)	(0.0065)	(0.0065)	(0.0065)	(0.0065)
Constant	0.0023*	0.0023*	0.0024*	0.0023*	0.0023*	0.0022*	0.0023*	0.0022*
	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)
CONLOC					0.0660**	0.0851***	0.0883***	0.0698***
					(0.0276)	(0.0178)	(0.0152)	(0.0133)
sq_CONTLOC					−0.0438*	−0.0720***	−0.0773***	−0.0689***
					(0.0242)	(0.0162)	(0.0140)	(0.0128)
DISCLOC_none	0.0001	−0.0037	−0.0057	−0.0129				
	(0.0139)	(0.0122)	(0.0118)	(0.0115)				
DISCLOC_double	0.0164**	0.0206***	0.0212***	0.0147***				
	(0.0080)	(0.0048)	(0.0039)	(0.0023)				
DISCLOC_center	0.0127	0.0095**	0.0087**	−0.0012				
	(0.0081)	(0.0049)	(0.0041)	(0.0037)				
PUB	0.0147***	0.0145***	0.0144***	0.0144***	0.0148***	0.0146***	0.0146***	0.0145***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
CIT_PUB	0.00004***	0.00004***	0.00004***	0.00004***	0.00004***	0.00004***	0.00004***	0.00004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
CARRY	0.0045***	0.0045***	0.0045***	0.0044***	0.0046***	0.0045***	0.0045***	0.0045***
	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)

Table 2 continued

	Model(1)	Model(2)	Model(3)	Model(4)	Model(5)	Model(6)	Model(7)	Model(8)
sq_CARRY	−0.0002***	−0.0002***	−0.0002***	−0.0002***	−0.0002***	−0.0002***	−0.0002***	−0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
N	80958	80958	80958	80958	80658	80658	80658	80658
Sum squared resid	3929.459	3927.627	3925.523	3926.008	3914.513	3913.312	3911.412	3912.287
S.E. of regression	0.2203	0.2203	0.2202	0.2202	0.2203	0.2203	0.2202	0.2202

$***p < 0.01; **p < 0.05; *p < 0.1$

Table 3 Regression results for ENQ

	Model(9)	Model(10)	Model(11)	Model(12)	Model(13)	Model(14)	Model(15)	Model(16)
NETPOS	log(ENQ)	log(ENQ)	log(ENQ)	log(ENQ)	log(ENQ)	log(ENQ)	log(ENQ)	log(ENQ)
GEOLOC	Discrete	Discrete	Discrete	Discrete	Continuous	Continuous	Continuous	Continuous
GEOCENT	Broad	EU15+	Narrow	USA	Broad	EU15+	Narrow	USA
PERF	Both	Both	Both	Both	Both	Both	Both	Both
NETPOS(−1)	0.6380***	0.6364***	0.6363***	0.6367***	0.6381***	0.06362***	0.6360***	0.6368***
	(0.0096)	(0.0096)	(0.0096)	(0.0096)	(0.0096)	(0.0096)	(0.0096)	(0.0096)
Constant	−0.0260***	−0.0252***	−0.0252***	−0.0253***	−0.0252***	−0.0243***	−0.0242***	−0.0245***
	(0.0060)	(0.0060)	(0.0060)	(0.0059)	(0.0059)	(0.0059)	(0.0059)	(0.0059)
CONLOC					0.3360***	0.3974**	0.4651***	0.3414***
					(0.1209)	(0.0785)	(0.0696)	(0.0553)
sq_CONTLOC					−0.3586***	−0.3392***	−0.4174***	−0.3743***
					(0.1021)	(0.0690)	(0.0623)	(0.0531)
DISCLOC_none	−0.0807	−0.0017	−0.0213	−0.0682				
	(0.0793)	(0.0731)	(0.0716)	(0.0707)				
DISCLOC_double	0.0656*	0.1019***	0.1116***	0.0670***				
	(0.0364)	(0.0218)	(0.0184)	(0.0135)				
DISCLOC_center	−0.0251	0.0585**	0.0445**	−0.0304*				
	(0.0379)	(0.0234)	(0.0207)	(0.0182)				
PUB	0.0637***	0.0633***	0.0625***	0.0622***	0.0640***	0.0635***	0.0629***	0.0628***
	(0.0018)	(0.0018)	(0.0018)	(0.0018)	(0.0018)	(0.0018)	(0.0018)	(0.0018)
CIT_PUB	0.0004***	0.0004***	0.0004***	0.0004***	0.0004***	0.0004***	0.0004***	0.0004***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
CARRY	0.0546***	0.0549***	0.0549***	0.0545***	0.0539***	0.0541***	0.0541***	0.0538***
	(0.0059)	(0.0059)	(0.0059)	(0.0059)	(0.0058)	(0.0058)	(0.0058)	(0.0058)

Table 3 continued

	Model(9)	Model(10)	Model(11)	Model(12)	Model(13)	Model(14)	Model(15)	Model(16)
sq_CARRY	0.0009***	0.0009***	0.0009***	0.0009***	0.0009***	0.0009***	0.0009***	0.0009***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
N	70624	70624	70624	70624	70394	70394	70394	70394
Sum squared resid	62982.38	62894.09	62875.71	62874.12	62717.25	62609.90	62574.37	62613.80
S.E. of regression	0.9441	0.9437	0.9435	0.9436	0.9440	0.9431	0.9429	0.9432

$* * * p < 0.01; * * p < 0.05; * p < 0.1$

The results presented in Tables 2 and 3 seem robust as the magnitude of the coefficients does not change considerably across the different model specifications and using different proxies for the concepts of our interest results in similar evidences.

5 Summary

An extensive literature confirms that scientific performance (which can be measured by publications or citations) significantly interrelates with knowledge network position (measurable by co-authorship networks). Related to this literature our paper focuses on the role geographical location of scientists plays in the development of their scientific network position. In order to do so we examined the co-evolution of geographical location, network position and scientific performance over the career of top researchers in neuroscience.

To measure network and geographical positions we developed a core-periphery categorization of authors in the network and geography dimensions. These dimensions relate to each other in a way that authors belonging to the network core tend to locate in the geographic center with a higher probability than locating in the geographical periphery. On the other hand, a remarkable core-periphery pattern exists in scientific linkages along the geographical dimension. Authors affiliated in the geographical center have almost exclusively links with other authors also from the geographical center, while the case is reversed for authors affiliated only in the periphery. Authors starting at the geographical periphery have very limited connections to the geographical center in the beginning of their career. They gain connections very rapidly to authors in the geographical center but they maintain their peripheral connections at the same time.

Despite better positions of the geographical core in new knowledge resources evidenced by a higher concentration of researchers belonging to the network core in the geographic center and more intensive scientific collaborations there we do not observe significant differences in publication activity between scientists located in the geographic core and researchers situated in the geographic periphery. On the other hand there are also no differences in favor of scientists located in the geographical core with respect to knowledge access measured by network quality and the share of links directing to the network core. It is thus suggested that research collaboration between the geographic periphery and the geographic core is the factor that helps peripheral regions maintain comparable scientific performance in terms of publication output.

However knowledge transfers mediated via co-authorship linkages between scientists of the geographical core and the periphery are not sufficient enough to develop comparable citation and network positions of researchers located in the periphery with those located in the geographic core. To make significant improvements in these respects more frequent linkages assured by physical proximity are in demand. The striking result of the analysis in this paper is that maintaining affiliations in both geographical core and periphery institutions acts as an efficient way to improve on publication, citation and network positions of researchers. Along these dimensions authors with double affiliations tend to outperform even those scientists who are affiliated only with the geographic core.

In this paper we presented our first results on the interrelations of geographic and network positions and scientific performance based on the longitudinal co-authorship data of neural scientists. In our future work we plan to enlarge our sample towards different scientific fields, develop more thorough author identification for critical authors as well as move the analysis directly to the link level.

Appendix

In this paper we use the formerly developed Ego network Quality (ENQ) index to measure knowledge accessible for a given actor in a network. In the Appendix we briefly describe the main concepts behind the measure. A more detailed introduction can be found in Sebestyén and Varga (2013a, b).

We structure the ENQ index around basically two dimensions, which are then augmented with a related third aspect. The two dimensions are: Knowledge Potential and Local Structure. Knowledge Potential (KP) measures knowledge accumulated in the direct neighbourhood and it is related to the number of partners and the knowledge of individual partners. Local Structure (LS) is associated with the structure of links among partners. The third aspect is called Global Embeddedness (GE) as it intends to capture the quality of distant parts of the network (beyond immediate partners). However, this aspect is implemented by applying the concepts of KP and LS for consecutive neighbourhoods of indirect partners in the network.

The notation in the proceeding formulation is as follows. We represent the network under question by the adjacency matrix $A = [a_{ij}]$, where the general element a_{ij} describes the connection between nodes i and j. The adjacency matrix defines the matrix of geodesic distancesâ (lengths of shortest paths) between all pairs of nodes, which is denoted by $R = [r_{ij}]$. In order to account for knowledge levels, we use $k = [k_i]$ as the vector of knowledge at each specific node of the network. By 'neighbourhood at distance d' we mean the nodes exactly at distance d from a specific node.

Given the conceptual model presented above, we formalize ENQ as follows:

$$\text{ENQ}^i = \sum_{d=1}^{M-1} W_d \text{LS}_d^i \text{KP}_d^i. \tag{A1}$$

In this formula superscript i refers to the node for which ENQ is calculated and subscript d stands for distances measured in the network (geodesic distance). M is the size of the network, W_d is a weighting factor used for discounting values at different d distances from node i, whereas KP_d^i and LS_d^i are the respective Knowledge Potential and Local Structure values evaluated for the neighbourhood at distance d from node i. The proposed formula for ENQ is a distance-weighted sum of Local Structure-weighted Knowledge Potentials evaluated for neighbourhoods at different distances

in the network. In what follows, the two basic concepts, Knowledge Potential and Local Structure are introduced in more detail.

Knowledge potential The concept of KP relates to the amount of knowledge an agent's partners possess. Using the notation presented before, the concept of KP can be formulated in the following way:

$$KP_d^i = \sum_{j:r_{ij}=d} K_j \qquad (A2)$$

The Knowledge Potential, as perceived by node i, can thus be calculated for the neighbourhoods at different d distances from node i, and for all these distances it is the sum of knowledge possessed by nodes at these distances.

Local structure Local Structure (LS), referring to the cohesion concept, is associated with the strength of ties and the intensity of interactions among partners. It is the sum of the tie weights present in a given neighbourhood, normalized by the size of this neighbourhood:

$$LC_d^i = \frac{1}{N_d^i} \left(\sum_{j:r_{ij}=d-1} \sum_{l:r_{il}=d} a_{jl} + \frac{\sum_{j:r_{ij}=d} \sum_{l:r_{il}=d} a_{jl}}{2} \right) \qquad (A3)$$

where N_d^i is the number of nodes laying exactly at distance d from node i. The expression in the parenthesis is made up of two parts. The first term counts the (weighted) ties between nodes at distance d − 1 and d[15] This reflects the intensity at which two adjacent neighbourhoods are linked together. The second term counts the (weighted) number of ties among nodes at distance d.[16] As a result, Local Connectivity captures the intensity with which the (possibly indirect) neighbours at distance d are linked together and linked to other neighbourhoods.

References

Abbasi, A., Jaafari, A.: Research impact and scholars' geographical diversity. J. Informetr. **7**(3), 683–692 (2013)

Abbasi, A., Altmann, J., Hossain, L.: Identifying the effects of co-authorship networks on the performance of scholars: a correlation and regression analysis of performance measures and social network analysis measures. J. Informetr. **5**, 594–607 (2011)

Abbasi, A., Chung, K.S.K., Hossain, L.: Egocentric analysis of co-authorship network structure, position and performance. Inf. Process. Manag. **4**(48), 671–679 (2012)

Acedo, F.J., Barroso, C., Casanueva, C., Galán, J.L.: Co-authorship in management and organizational studies: an empirical and network analysis. J. Manag. Stud. **43**(5), 957–983, 07 (2006)

Agrawal, A., Cockburn, I., McHale, J.: Gone but not forgotten: labour flows, knowledge spillovers, and enduring social capital. J. Econ. Geogr. **6**, 571–91 (2006)

Ahn, J., Oh, D., Lee, J.: The scientific impact and partner selection in collaborative research at Korean universities. Scientometrics **100**(1), 173–188 (2014)

---

[15] Distances are always measured from node *i*.

[16] Division by two is required because matrix A is symmetric, and thus we can avoid duplications in the counting.

Autant-Bernard, C., Billand, P., Frachisse, D., Massard, N.: Social distance versus spatial distance in R&D cooperation: empirical evidence from European collaboration choices in micro and nanotechnologies. Pap. Reg. Sci. **86**, 495–519 (2007)

Barabasi, A.L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the social network of scientific collaborations. Physica A **311**(3–4), 590–614 (2002)

Boschma, R.: Proximity and innovation: a critical assessment. Reg. Stud. **39**, 61–74 (2005)

Braczyk, H., Cooke, P., Heidenreich, H.: Regional Innovation Systems: The Role of Governances in a Globalized World. UCL Press, London (1998)

Broekel, T., Boschma, R.: Knowledge networks in the Dutch aviation industry: the proximity paradox. J. Econ. Geogr. **12**, 409–433 (2012)

Burt, R.S.: Structural Holes. Harvard University Press, Cambridge (1992)

Cunningham, S.W., Werker, C.: Proximity and collaboration in European nanotechnology. Pap. Reg. Sci. **91**, 723–742 (2012)

Della Rossa, F., Dercole, F., Piccardi, C.: Profiling core-periphery network structure by random walkers. Sci. Rep. **3**, 1467 (2013)

Fafchamps, M., van der Leij, M.J., Goyal, S.: Matching and network effects. J. Eur. Econ. Assoc. **8**, 203–231 (2010)

González-Pereira, B., Guerrero-Bote, V.P., Moya-Anegón, F.: A new approach to the metric of journals' scientific prestige: the SJR indicator. J. Informetr. **4**(3), 379–391 (2010)

Hansen, Teis: Substitution or overlap? The relations between geographical and non-spatial proximity dimensions in collaborative innovation projects. Reg. Stud. **49**(10), 1–13 (2014)

Hardeman, S., Frenken, K., Nomaler, Ö., ter Wal, A.: A proximity approach to territorial science systems. Paper prepared for the EUROLIO Conference on "Geography of Innovation" Saint-Etienne, France, 24–26 January 2012 (2012)

Hazir, C., Autant-Bernard, C.: The effect of spatio-temporal knowledge flows on regional innovation performance: the case of ICT patenting in Europe. SEARCH working paper (2013)

Jeong, S., Choi, J.Y., Kim, J.: The determinants of research collaboration modes: exploring the effects of research and researcher characteristics on co-authorship. Scientometrics **89**, 967–983 (2011)

Liu, X., Bollen, J., Nelson, M.L., Van de Sompel, H.: Co-authorship networks in the digital library research community. Inf. Process. Manag. **41**, 1462–1480 (2005)

Lundvall, B. (ed.): National Systems of Innovation. Toward a Theory of Innovation and Interactive Learning. Anthem Press, London (2010)

Maggioni, M., Nosvelli, M., Uberti, E.: Space vs. networks in the geography of innovation: a European analysis. Pap. Reg. Sci. **86**, 471–493 (2006)

Nelson, R. (Ed.): National innovation systems. Oxford, New York (1992)

Newman, M.E.J.: Coauthorship networks and patterns of scientific collaboration. Proc. Natl. Acad. Sci. USA **101**, 5200–5205 (2004)

Rodriguez, M.A., Pepe, A.: On the relationship between the structural and socioacademic communities of a coauthorship network. J. Informetr. **2**(3), 195–201 (2008)

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

Sebestyén, T., Varga, A.: Research productivity and the quality of interregional knowledge networks. Ann. Reg. Sci. **50**(1), 155–189 (2013a)

Sebestyén, T., Varga, A.: A novel comprehensive index of network position and node characteristics in knowledge networks: ego network quality. In: Scherngell, T. (ed.) The Geography of Networks and R&D Collaborations, pp. 71–97. Springer, New York (2013b)

Varga, A. (ed.): Universities, Knowledge Transfer and Regional Development: Geography, Entrepreneurship and Policy. Edward Elgar Publishers, Northampton (2009)

Varga, A., Sebestyén, T.: Does EU Framework Program participation affect regional innovation? The differentiating role of economic development. Int. Reg. Sci. Rev. (2017) (forthcoming)

Varga, A., Pontikakis, D., Chorafakis, D.: Metropolitan Edison and cosmopolitan Pasteur? Agglomeration and interregional research network effects on European R&D productivity. J. Econ. Geogr. **14**, 229–263 (2014)

Yan, E., Ding, Y.: Applying centrality measures to impact analysis: A coauthorship network analysis. J. Am. Soc. Inf. Sci. Technol. **60**(10), 2107–2118 (2009)