Global research community: what do international co-authorship patterns say about its structure? *

Comunidad de investigación global: ¿qué dicen los patrones de coautoria internacional sobre su estructura?

Comunidade de pesquisa global: o que os padrões internacionais de coautoria dizem sobre sua estrutura?

Received: 20 de septiembre de 2018. Accepted: 11 de octubre de 2018

Written by:
S.G. Shulgin\textsuperscript{31} 
Ju.V. Zinkina\textsuperscript{32} 
A.I. Andreev\textsuperscript{33} 
A.T. Gasparishvili\textsuperscript{34}

Abstract

Global research community represents a noteworthy object for sociological studies. In this paper we regards it as an entity with network structure and try to analyze the current state of this structure and its recent dynamics. Among the many various approaches to studying the network links that bring together the researchers all over the world, we choose to look at the structure of international co-authorship of research papers published in journals indexed by Scopus Elsevier database. We try to provide a general outline of the network structure, to reveal distinct regional clusters existing at the various levels of this structure, and to discover whatever significant changes took place in the network structure over the last 25 years.

Keywords: International co-authorship, international co-authorship network, the USA, the UK, China, global research community.

Resumen

La comunidad de investigación global representa un objeto notable para los estudios sociológicos. En este documento, lo consideramos como una entidad con estructura de red e intentamos analizar el estado actual de esta estructura y su dinámica reciente. Entre los diversos enfoques para estudiar los enlaces de red que reúnen a los investigadores de todo el mundo, elegimos observar la estructura de la coautoria internacional de los trabajos de investigación publicados en revistas indexadas por la base de datos Scopus Elsevier. Intentamos proporcionar una descripción general de la estructura de la red, revelar los distintos grupos regionales que existen en los distintos niveles de esta estructura y descubrir los cambios significativos que tuvieron lugar en la estructura de la red durante los últimos 25 años.

Palabras claves: Coautoria internacional, red de coautorial internacional, EE.UU., Reino Unido, China, comunidad de investigación global.

\textsuperscript{31} Candidate of Economic Sciences, Assistant Manager of the International Laboratory of Demography and Human Capital of RANEPA, Email: global@ores.su
\textsuperscript{32} Candidate of Historical Sciences, Senior Research Officer of the International Laboratory of Demography and Human Capital of RANEPA, Email: info@ores.su
\textsuperscript{33} Candidate of Biology, Associate Professor of Faculty of Global Processes of Lomonosov Moscow State University, Email: russia@prescopus.com
\textsuperscript{34} Candidate of Philosophical Sciences, Associate Professor of Faculty of Global Processes of Lomonosov Moscow State University, Email: editor@ores.su

*"The article was prepared within the framework of the project (unique identification number RFMEFI57217X0005) with the financial support of the Ministry of Education and Science of the Russian Federation". 

Encuentre este artículo en http://www.udla.edu.co/revistas/index.php/amazonia-investiga ISSN 2322- 6307
Resumo

A comunidade global de pesquisa representa um objeto notável para os estudos sociológicos. Neste trabalho nós o consideramos como uma entidade com estrutura de rede e tentamos analisar o estado atual dessa estrutura e sua dinâmica recente. Entre as várias abordagens para o estudo dos links de rede que reúnem pesquisadores de todo o mundo, optamos por analisar a estrutura de coautoria internacional de trabalhos de pesquisa publicados em periódicos indexados pela base de dados Scopus Elsevier. Tentamos fornecer um esboço geral da estrutura da rede, revelar clusters regionais distintos existentes nos vários níveis dessa estrutura e descobrir as mudanças significativas ocorridas na estrutura da rede nos últimos 25 anos.

Palavras-chave: Coautoria internacional, rede internacional de coautoria, EUA, Reino Unido, China, comunidade global de pesquisa.

Introduction

One of the most essential aspects of globalization is related to increased global connectivity, and that can be found in numerous spheres of human world, including science and research. Various types of connections bond the global community of researchers together, and one of the most important ones is cooperated work on a certain research problem that results in producing co-authored research output. Drawing on the expertise of fellow scholars to achieve higher-quality research is the very essence of cooperation in scientific world. Be it the discussion of papers at conferences, peer-review procedures run by the majority of journals, of informal asking a colleague researcher for opinion/comments, cooperation is undoubtedly a tenet of the world of science. There is ample evidence that international scientific collaboration has been increasing in the recent decades (Wagner, et al. 2015). This increase is largely related to more general processes of global integration. Say, the European Union has been actively promoting scientific collaboration between its member countries; indeed, many large-scale research projects could only be implemented at a regional level rather than within the borders of a single country (Glänzel, et al. 1999).

There are multiple ways to measure such cooperation at the international level (e.g., through the number and scale of international conferences, bilateral or multilateral research programs, joint patents, number of foreigners holding research positions in a given country’s research and higher educational institutions, etc.). One study focusing on the role of EU countries as scientific collaborators of other countries measures the relative citation impact of internationally co-authored publications and reveals that “international scientific collaboration is particularly advantageous for less advanced countries, but also highly industrialised countries benefit from collaboration” (Glänzel, et al. 1999). This result is corroborated by another study focused on Eastern Europe. This study shows that international collaboration strongly intensified in the Eastern European academic communities during the 1990s and the 2000s, and “scientists from developed countries within the European Union play a key role in stimulating the international collaboration of academics in this region” (Teodorescu & Andrei 2011). Another piece of research investigates Latin American scientific output, analyzing the scientific cooperation patterns of LA countries with several geographic regions: amongst themselves, with the EU and with the USA. The analysis reveals that collaboration patterns of LA countries are heterogeneous; e.g. the size of the country is inversely related to its international collaboration rate (Gomez, et al. 1999). Data from Elsevier Scopus database was used to assess the scale of scientific collaboration among the BRICS countries (Brazil, Russia, India, China, and South Africa). A number of BRICS official documents declare the intention of developing scientific cooperation. The empirical evidence on the actually existing cooperation between the BRICS, however, is mixed: “in the group of BRICS countries some relatively strong scientific collaboration ties exist. Nevertheless, such ties are not necessarily the strongest the country experiences. Moreover, data on the evolution of the collaborations show that, if a “BRICS effect” on the strength of collaborations exists, this is not so strong” (Finardi 2015). Within the BRICS, China represents a particularly interesting case for scientific collaboration studies, as this country has recently been making extreme progress in its
stance and visibility in the international scientific arena. A thorough study of Chinese provinces’ participation in international scientific collaboration has revealed “different collaborative patterns corresponding to development levels in economy, science and technology” (Wang et al. 2005). Collaborative patterns also differ in terms of subjects are related to the characteristics of the subject itself; hierarchical cluster analysis has been revealed clinical medicine to have the highest absolute number and relative proportion of co-authored papers (Wang et al. 2005).

However, most of the existing research concentrates on some particular regions or countries. Of greater interest to us is the systemic view of international scientific cooperation at the global level into which all the abovementioned regional structures and patterns fit. In this paper, we choose to measure international cooperation in various fields of research through co-authorship of research papers published in journals indexed by Scopus, the largest abstract and citation database of peer-reviewed literature run by Elsevier. Our aim is to answer a number of questions about the structure of international scientific cooperation, namely:

1) Does this structure have a distinct center-periphery structure (a group of intensely cooperating countries against the background of a larger number of less connected ones)?
2) Are there any distinct regional clusters within this structure?
3) Has this structure experienced any significant changes in the recent decades and, if so, what were the likely causes of these changes?

The structure of the paper is as follows. Section 2 describes data and methods used in our research. Section 3 provides some of the most notable results and their discussion. Section 4 concludes with a view to getting the answers to the three research questions listed above.

Data and Methods

Dataset for this research has been retrieved from the Elsevier Scopus database of peer-reviewed literature, currently the largest existing database of the kind (Elsevier 2018a). It includes over 23,700 peer-reviewed journals covering numerous branches of research from health sciences, physical sciences, social sciences, and life sciences. These journals originate from various regions of the world and are supplemented by more than 166,000 books and over 8.3 million conference papers. Along with some other types of publications, all these total to over 71 million records overall (and over 1.4 billion cited references) (Elsevier 2018b). The structure of the database is very useful for obtaining the scientometric data on international scientific collaborations in the form of international co-authorship of research papers.

We collected our data in January-April 2018, focusing on the period from 1991 to 2016. The year 1991 was specified as the lower limit of our time span because earlier data are more likely to be incomplete (the deeper in time, the more so) and to account only for a fraction of publications produced by major publishers.

In our analysis we rely on several types of data collected from Scopus database and aggregated by years (for each year from 1991 to 2016). Descriptive statistics for all these types of data are presented in Table 1.

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of countries reported</th>
<th>Number of countries mentioned</th>
<th>Country-to-country data points</th>
<th>Year</th>
<th>Number of countries reported</th>
<th>Number of countries mentioned</th>
<th>Country-to-country data points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1991</td>
<td>134</td>
<td>191</td>
<td>3,991</td>
<td>2004</td>
<td>142</td>
<td>217</td>
<td>10,748</td>
</tr>
<tr>
<td>1992</td>
<td>135</td>
<td>197</td>
<td>4,564</td>
<td>2005</td>
<td>143</td>
<td>216</td>
<td>10,767</td>
</tr>
</tbody>
</table>

Table. 1. Structure of SCOPUS publication database
First column (1) shows the year to which the data presented in a given row belongs.

Second column (2) shows the total number of unique “base” countries in a given year. For each “base” country we collected statistics on the number of its co-authorships with other countries. In each case we used Scopus report for a certain country for a given year that shows the number of cross-country co-authorships entered by this country in the given year.

Third column (3) shows the total number of “referenced” countries – i.e., the number of unique countries mentioned in a given year in all country reports as participants of of cross-country co-authorships with “base” countries (Kibalnik & Volosyuk, 2017).

Forth column (4) shows for a given year how many bilateral country-to-country data points we collected. Each data point indicates the presence in Scopus database of at least one paper co-authored by researchers from a given pair of countries in the given year. For example, a data point for Germany and Russia in the year 2000 would mean that Scopus indexed at least one (or more) paper co-authored by Russian and German scholars in this year.

We view the database of bilateral country-to-country data points as a network of nodes (represented by countries) and links between them (represented by the presence of co-authorship between a given pair of countries). The methodology used to analyze this network is based on the theory of graphs. We build a graph where each country corresponds to a vertex, and the edge between two vertices of the graph corresponds to the number of publications co-authored by researchers from these two countries. However, accounting for all the existing international co-authorships would create a very dense graph with numerous edges, which would complicate further analysis of the structure of the network and understanding the dynamics of international cooperation. Therefore, we choose to visualize only the structure of main co-authorship partners – that is, we include only one edge for each country, and it would be the “thickest” edge denoting the link with this country’s most important partner in terms of international co-authorship (Lankina & Platonova, 2015).

In other words, if a country A has co-authored papers with researchers from 27 countries, we do not put 27 edges on the graph, but only one – linking it to country B, with which country A shares the largest number of co-authored papers.

**Results**

We present a visualization of the “main co-authorship partners” graph on Figure 1. The area (size) of each vertex is proportionate to the total amount of publications affiliated with this country. Edge width is proportionate to the number of publication publications co-authored by this pair of countries (log-scale). Name of each
vertex is the ISO alpha-3 country code. Also, we use regional color coding to represent countries from the same region with the same color and countries from the same continent (or the same part of the world) with similar colors. E.g. all African countries in are presented in red colors; European in green; Northern, Southern
and Central Americas (also Caribbean) in different tones of blue, etc. Major partner graph is a directed graph by its nature and the edge direction is also color-coded – the color of the edge corresponds to the color of vertex where this edge starts.

Figure 1 reveals the structure of the global network of main co-authorship partners to be of a rather distinct “hub-and-spoke” type. There is a clear hub, represented by the USA, which is the main co-authorship partner for the absolute majority of countries in various regions of the world. This means that for most countries, out of all internationally co-authored papers they have, the largest share is co-authored with researchers from the USA. There are some exceptions – for example, Germany happens to be the main co-authorship partner for Austria, Russia plays this role for a number of post-Soviet countries (such as Belarus, Kazakhstan, and a number of others); a certain cluster is formed around France, mainly including some of the

---

French-speaking African countries. However, the general predominance of the USA in international scientific cooperation (reflected in paper co-authorship) remains indisputable.

In order to disclose deeper structural layers of this network, let us try to remove the USA from the list of partners (but still have it as a vertex in the network) and investigate the structure of the remaining relations (this means that for the countries having the USA as their main partner, we would switch to their 2nd largest partner). This new network is visualized in Fig. 2.

Fig. 2 reveals a significantly different network structure, as a number of distinct clusters replace the “hub-and-spoke” structure present in Fig. 1. We can see that United Kingdom plays the role of the Global Vice-Leader, being the 2nd largest co-authorship partner for many European, Middle Eastern, and African countries. However, a much clearer structure of regional scientific cooperation emerges in Fig. 2 as compared to Fig. 1. Indeed, in Fig. 2 the UK serves as the major “hub”, but it is not display such an extent of global predominance in the world of scientific cooperation as the USA does. A number of regional clusters appear along with the cluster centered on UK – most notably, two Latin American clusters, one attracted to Spain and another (smaller) one to Brazil. Two East Asian clusters centered on China and Japan merge into a single sizeable cluster. More countries attach themselves to the European cluster gravitating towards Germany (among the members of this cluster there is Russia, which serves as the main partner for its own sub-cluster). France preserves its own cluster consisting mostly of Francophonic African countries. Still, all the centers of these clusters have UK as their main partner in scientific cooperation, so we can say that UK holds together almost all the global scientific cooperation network.
What happens if we discard all the countries’ links with both the USA and the UK, setting for their next largest partner in co-authorships? Again, we use the same technique excluding edges to UK from the graph layout. At the next step we create a graph visualization for main co-authors (excluding all the links to USA & UK, but still keeping them as separate nodes in the network) – see Fig. 3.

After excluding from the visualization the two globally dominant countries, we could see in details the patterns of regional cooperation. Two major clusters are revealed – the European one (almost all European countries have Germany as their major co-authorship partner in this layout), and the Asian one, centered on China. Importantly, after the removal of the links to Great Britain, Canada and Australia shift to the Asian cluster. A number of smaller clusters are visible along with these two major clusters – some do not experience much change from the previous layout (the clusters centered on France and Russia), but some begin to look markedly different. Thus, Spain emerges as the major co-authorship partner for the majority of Latin American countries (for some countries of the region, Brazil or Mexico play this role, but the cluster as a whole is clearly centered on Spain). Russia and France preserve their role of sub-centers, serving as the major partners in scientific collaboration for post-Soviet and Francophonic African countries accordingly. Notably, another African sub-cluster centered on South Africa emerges along with the French-centered one. This newly emerged sub-cluster is also attracted towards the European cluster, as Germany is the largest scientific cooperation partner for South Africa.
If we proceed to remove all the links to Germany from the network, its structure experiences some further changes. The European cluster falls apart into two clusters centered on France and Italy. Removing the links to two more central nodes represented by two European countries, France and Spain, allows us to discover the rather surprisingly central role of Italy, which starts to hold together the whole of the European cluster. The role of Spain as the center of Latin American cluster is taken by Brazil instead. Let us now look at the deeper structural layers of the Asian cluster. China is its major center holding the cluster together; if we proceed to remove all the links to China, the outlook of the cluster change remarkably. Japan emerges as the “secondary” center, serving as the major co-authorship partner for a (much smaller) group of Asian countries. Some more countries get attached to Australia and Malaysia (see Fig. 4).

Finally, if we proceed to remove the links to all the central nodes (the USA, Great Britain, Germany, France, Spain, Italy, and China), the network of major partners in international co-authorship completely falls apart into a number of separate clusters which are mostly region-based and do not have any links to each other. The European cluster persists, but with no distinct central node. Eastern Europe forms a separate cluster along with some post-Soviet countries. Latin American cluster stays separate with Brazil and Mexico as its most prominent nodes. A rather pronounced cluster of Middle Eastern and North African countries forms around Saudi Arabia and, to a lesser extent, Egypt. The majority of Sub-Saharan countries have their major partnerships with South Africa. There is an Asian cluster, whose largest nodes are Japan, South Korea, and India. Finally, there is one “mixed” cluster held together by Canada and Australia. The latter serves as the major co-authorship partner mostly for some Asian countries, while the former, rather surprisingly, plays the same role for some Middle Eastern countries. Apart from this rather small Canadian cluster, almost all other links stay within borders of geographical/historical regions.

Fig. 4. Global network of main co-authorship partners (excluding the USA, the UK, Germany, France, and China), 2016
Let us now turn to another important question – how stable have these two major clusters been in time, and have they experienced any significant changes regarding their internal structure? To answer this question, we have built visualizations for year 1991 (see Figs 5-8 accordingly) using the same methodology that we used for visualization in Figs 1-4 (described above).

When comparing this structure with the one 25 years later (see Fig. 1) we notice that over this period the USA managed to preserve its role of the globally dominant international co-authorship partner for countries from various regions of the world. However, a number of changes are also noticeable. In 1991, Great Britain used to have many more countries in its cluster than 25 years later, while the opposite is true for Germany, which considerably expanded its influence in Europe over the same period. No Asian sub-centers were visible in 1991, and the absolute majority of Asian countries had the USA as their most important international co-authorship partner. This gives us grounds to expect significant difference at deeper structural layers of the network as well. To reveal them, we will proceed to remove the links to the most central nodes from the network in the same fashion as we did for the 2016 structure. We start by removing all the links to the USA; the resulting network structure is present in Fig. 6.
Some notable changes are visible at this structural level that took place between 1991 and 2016 (compare Fig. 9 with Fig. 2). Great Britain preserved its role of the 2nd leader in the international co-authorship network, but in 1991 it was the main partner for many more Asian countries than in 2016. Germany persisted in its role of the major partner for many European countries. Other clusters experienced greater change. France, for example, lost its centrality for many Latin American countries to the newly emerged central nodes of Spain and, to a lesser extent, Brazil. East Asian cluster grew in size and had its center shift from Japan to China (though Japan retained a sub-cluster of its own). Moreover, a number of smaller centers emerged by 2016 (e.g., India, Malaysia, South Africa, Brazil, Saudi Arabia), bringing greater variety into the network structure.

Proceeding to remove all the links to Great Britain, we discover the dynamics of an even deeper structural layer (Fig. 7).
In 1991, France and Germany shared the 3rd place in the ranking of most important international co-authorship partners for countries worldwide (after the USA and the UK). Germany played this role for many European countries and quite a number of Asian ones. France mostly attracted a variety of Latin American and African countries. These two major clusters represented the majority of the world counties; some sub-clusters were present as well (with such centers as Australia, Japan, Russia, Sweden, or Canada), but they were much smaller and generally integrated into one of the two major clusters. By 2016, this structural level of network changed to display much greater variety. First of all, France clearly lost a great part of its power, and could not claim the role of a globally dominant co-authorship partner any longer. Indeed, in 2016 its influence on the network of the major international co-authorship partners was mostly limited to Francophone African countries. Germany, on the contrast, retained its position and even strengthened its role, “gathering” almost all European countries in its cluster. Second, the global research community was no longer divided into just two groups, as a number of new clusters with newly emerged centers appeared in the network. Thus, Spain rose to play a prominent role of the major co-authorship partner for the majority of Latin American countries. South Africa began to rival France in the scale of its influence on African countries, becoming the major co-authorship partner for quite a number of English-speaking African countries. And, most notably, a huge Asian cluster emerged, centered on China.

Finally, if we remove all the links to Germany and France, we reveal some more noteworthy patterns in the structure of international co-authorship network in 1991 (see Fig. 8). The division of the world into two major clusters would persist, but each of these clusters would have a few different centers rather than a single one. The major non-European partners in international co-authorship were represented by Japan, Canada, and Australia.
Conclusion

The analysis of the global network of main partners in scientific cooperation yields a number of noteworthy results discovering rather interesting patterns. First of all, the general structure of this network appears to be of “hub-and-spoke” type, with the USA rather predictably playing the role of the central hub (it persisted in this role for at least 25 years). Once all the links to the USA are removed from the network, the UK takes the place of the central hub, but the structure of the network changes remarkably – it is not “hub-and-spoke” anymore, as a number of distinct regional clusters with their respective centers appear. However, the UK can still be said to be the 2nd globally dominant scientific cooperation partner, as it happens to be the main partner for all the centers of the regional clusters, thus holding the whole network together. As we proceed to remove all the links to the UK from the network, its structure again changes remarkably, as two distinct clusters appear, the European one (centered on Germany), and the Asian one (centered on China). Each cluster has its own sub-clusters, but all of those are held together by their ties to Germany or China respectively. However, the co-existence of these two clusters is a rather recent phenomenon, as in 1991 there was no sign of a separate Asian cluster, and countries of the world were divided between the German and the French clusters at this structural level of the network. In general, removing the influence of the globally dominant international co-authorship partners allows revealing notable patterns of smaller clusters emerging around new centers. Thus, Italy has increased its influence as a co-authorship partner among the European countries, while Spain took the role of the major partner for the majority of Latin American countries. The number of such clusters
was smaller in 1991, and they were mostly centered on developed countries. By 2016, on the contrast, such developing countries as, say, Brazil or South Africa emerged as major co-authorship partners for quite a number of countries in their respective regions.

So, what does the network of major partners in international co-authorship tell us about the structure of the global community? At the top level, the USA retains its dominant influence. However, at deeper structural layers, significant changes have been taking place, with some countries losing their prominent positions (France), and others rapidly rising (China). The structure of the network has been gaining variety, as the list of major partners in international co-authorship is longer now, and includes both developed and developing countries (the impact of the latter is limited yet, but has clearly been on the rise over the last decades). However, while at the global level the links of various countries to the USA have permeated the borders of the regions, at deeper structural levels the regional belonging still exerts critical influence over the network structure. Most of the clusters correspond to certain regions, defined by geographical proximity and/or common language and/or shared historical past. Apart from the USA and the UK, almost no country managed to become a major co-authorship partner for any significant number of countries outside its geographically or historically defined region.

However, our analysis provides sufficient evidence for the rapid and powerful “rise of the East” in international scientific cooperation in the last quarter of a century. China already rivals Germany; the dominance of the USA and the 2nd leading role of the UK) remains indisputable so far, but further strengthening of the Chinese cluster can bring further structural changes into the network.

References

Elsevier. 2018a. Scopus Database. URL: https://www.scopus.com