Networks of international collaboration and mobility: a comparative study

Zaida Chinchilla-Rodríguez[1], Lili Miao[2], Dakota Murray[2], Nicolás Robinson-García[3], Rodrigo Costas[4], and Cassidy R. Sugimoto[2, 4]

[1] *zaida.chinchilla@csic.es*
Consejo Superior de Investigaciones Científicas (CSIC) (IPP), SCImago Group (Spain)

[2] *lilymiao08@gmail.com; dakota.s.murray@gmail.com; sugimoto@indiana.edu*
Indiana University Bloomington (USA)

[3] *elrobinster@gmail.com*
Universitat Politècnica de València (INGENIO-CSIC-UPV), Valencia (Spain)

[4] *rcostas@cwts.leidenuniv.nl*
Centre for Science and Technology Studies (CWTS), Leiden University, Leiden (Netherlands)

Abstract
This study presents a preliminary comparison of networks of international collaboration and mobility. Using affiliation data from scientific publications, we analyse the structural differences in the two networks and the role of countries. The results show that researchers collaborate internationally to a much higher degree than they become internationally mobile. The number of countries involved in the networks is three times higher in collaboration than in mobility, and the average degree demonstrates that mobility networks form tight structures with fewer links than collaboration networks. The role of countries differs between the collaboration and mobility network, predominately reflecting income level. Limitations and future research are described to further understand the dynamics of collaboration and mobility networks.

Conference Topic
Scholarly Communication; scientific mobility; collaboration; network analysis

Introduction
There is a popular saying: “Science knows no borders”. Indeed, scientists often move and collaborate internationally, accessing “multiple learning environments”, usually producing higher-impact research, and increasing their prestige (Gazni, et al. 2012; Glänzel 2001; Chinchilla et al. 2012). There are many well-documented reasons why researchers collaborate (e.g., Beaver 2001; Sonnenwald, 2007). Previous studies focus on the factors driving collaboration between countries, such as their relative size, their geographical, historical, linguistic, and thematic proximity, or their socio-economic characteristics (Zitt et al. 2000; Adams 2014).

Mobility has been advocated as key to increasing the efficiency and effectiveness of research (OECD 2008, 2010; Scellato et al., 2015). While some studies focused on the economic and development impact caused by mobility (Gibson & McKenzie 2012), few have utilized bibliometric measures to understand scientific mobility (Moed & Halevi, 2014; Sugimoto, Robinson-Garcia, & Costas, 2016). Furthermore, no study to-date has compared, at a global scale, the differences in the scientific networks created through collaboration and mobility. This is necessary to understand whether mobility provides a novel or a duplicative lens on the network created through scientific collaboration.

We present here, using collaboration and mobility data drawn from contemporary bibliometric data (2008-2015), the comparative network statistics and ranking for more than 200 countries considering socioeconomic indicators and grouping according to scientific and technological capacities. These data can inform the understanding of and relationship between indicators of international collaboration and mobility.

Data and Methods

Data were retrieved from Web of Science database for the period 2008-2015, considering only authors that have at least two publications and for whom this period represents the date of first publication. A total of 3,521,797 individual's authors were identified using a large-scale disambiguation algorithm developed by Caron and van Eck (2014). We consider all document types for the analysis. The nation-to-nation links were collected based on 14,097,939 publications. Country affiliations were extracted and cleaned, resulting in 213 unique countries. Collaboration linkages are measured considering all authors of each paper and their connections to countries. Mobility includes both migration (i.e., change from one affiliation to another) and co-affiliation (having multiple affiliations on the same publication or on two or more publications in a given year). Income level group was added as a proxy of wealth intensity of countries (World Bank 2016) and the Scientific and Technological Capacity Index—which measures expenditure in R&D, number of researchers, number of institutions, patents--(Wagner et al., 2001) was used as a proxy for the scientific capacities of countries.

For each group, we created and analyzed the subset of documents and researchers by calculating: the total number of publications per country and the number of publications with international collaboration. We use these two indicators as input to the international collaboration network. We build the international mobility network using the total number of active researchers per country and the number of active researchers linked to countries. We used integer counting, attributing a count of "1" to each occurrence of authorship or affiliation from a country.

Once we obtained the co-occurrence frequencies, we generated symmetric matrices using Pajek 64 4.10 (Batagelj & Mrvar, 1997) to derive statistical properties and structure of the networks. We explore the position and role of countries within each group using classical measures of network analysis. Closeness is calculated as the reciprocal of the sum of the shortest path between the nodes to other nodes (Freeman, 1978). Betweenness for a node is defined as the fraction of shortest paths among the whole network that pass through the node (Freeman, 1977). Density determines the degree of cohesion that exists among the nodes, revealing whether the network has a thick or thin consistency (Wasserman & Faust, 1999). The average degree measures the spread of influence across the networks (Hanneman & Riddle, 2006). Diameter measures the longest distance between a pair of countries, (i.e., how many steps from any node are necessary to reach any other node in the network (De Nooy et al., 2011)). The clustering coefficient indicates the proportion between the number of links in the neighborhood of a node and the number of links possible in the entire network (Watts & Strogatz, 1998; Barábasi, 2002). Assortativity is a property that reflects the tendency of nodes to connect to other nodes to a similar degree by means of measuring the Pearson correlation coefficient of degree between pairs of linked nodes (Newman, 2002).

For the visualization network, the proportions of collaborations and co-affiliations among countries were used as a pair of inclusion indexes (share of county j in the total co-authorships/co-affiliations of i or Affinity Index AFI (i,j), and its counterpart AFI (j,i) to characterize asymmetrical relationships between two countries (Zitt et al., 2000). AFI is a measure of the amount of collaboration or co-affiliations between a given country A and another country B, compared to the total collaboration or co-affiliation of the country A and country B. To position the countries (nodes) we applied the Kamada-Kawai algorithm (1989), characterized by its ability to assign coordinates to the vertexes while adjusting the distances to a maximum with respect to theoretical distances. Lines indicate the relations among actors, with the color (on a scale of grey tones) indicating the intensity of the connection. The colors of countries refer to geographical regions.

Results

Each network, summarized in Table 1, has 212 countries, one less than the original 213 because South Sudan is isolated in the collaboration network and Tuvalu is isolated in mobility. Differences in the number of edges, the density, and the average degree of each network suggests significant differences in their structural cohesion. The collaboration network has a higher density than the mobility network. Nodes in the collaboration network have three times the average degree than nodes in the mobility network. In comparison, the average shortest distance varies only slightly, indicating that the two networks are cohesively connected, where countries could reach each other in an average of less than two steps. However, considering their diameter, a distinctive feature appears: the collaboration network has a diameter of two, meaning that it is tightly linked with short distances between any pair of nodes; however, in the mobility network the diameter is twice as large, suggesting that the two networks are structurally different. The clustering coefficient is high in both networks demonstrating that countries form tightly knit groups, but the coefficient is slightly higher in collaboration than in mobility, indicating that two countries who have common neighbors would be slightly more likely to collaborate than have mobile researchers in common. However, considering that there are more than 2.5 times as many links in collaboration as in the mobility network; the mobility network has apparently formed tighter structures with fewer links. The negative values of assortativity indicate that countries with a large degree tend to connect with countries with a small degree in both networks.

Table 1. A summary of the properties of the two networks.

	Collaboration	*Mobility*
Number of nodes	212	212
Number of edges	12596	4788
Density	0.55	0.21
Average degree	118.27	44.95
Average shortest distance	1.44	1.84
Diameter	2	4
Cluster Coefficient	0.83	0.73
Assortativity	-0.19	-0.34

The rank of countries in collaboration and mobility networks

Figures 1-4 show the position of countries in closeness (left), betweenness (center) and clustering coefficient (right) for each network: with collaboration in the left panel and mobility in the right. We show four groups of countries according to their scientific and technological capacities where the colors of lines represent the income levels of countries: high (green), upper-middle (blue), lower middle (orange) and low-income (red). The interpretation of ranking labels goes as follows: Australia (3) (5) means that it ranks in the third position in the collaboration network and in the fifth position in the mobility network.

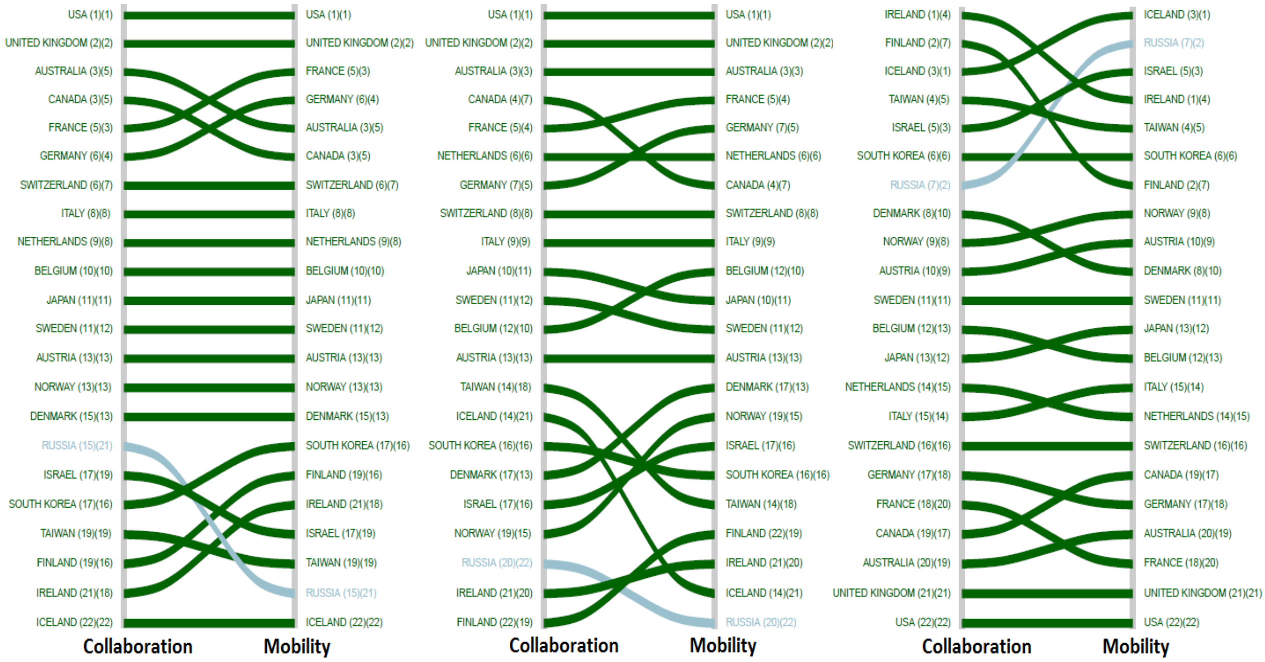

Figure 1. Closeness, betweenness and clustering coefficient for scientifically advanced countries.

In the group of scientifically advanced countries (Figure 1), a county's position is relatively stable across the two networks, suggesting that there exists a relatively similar ability to attract collaboration and mobility partners. The most striking situation is in Russia (upper-middle income) ranked relatively low in closeness and betweenness, whereas it has the second highest clustering coefficient in the mobility network; this suggests that Russia is not located in a central position, instead forming a tight group of partners, especially in mobility. While France, Germany, Denmark, Finland, and Ireland have more capacities to reach other countries and act as bridges especially in the mobility network, Canada, Taiwan, and Iceland show a more important role in the collaboration network.

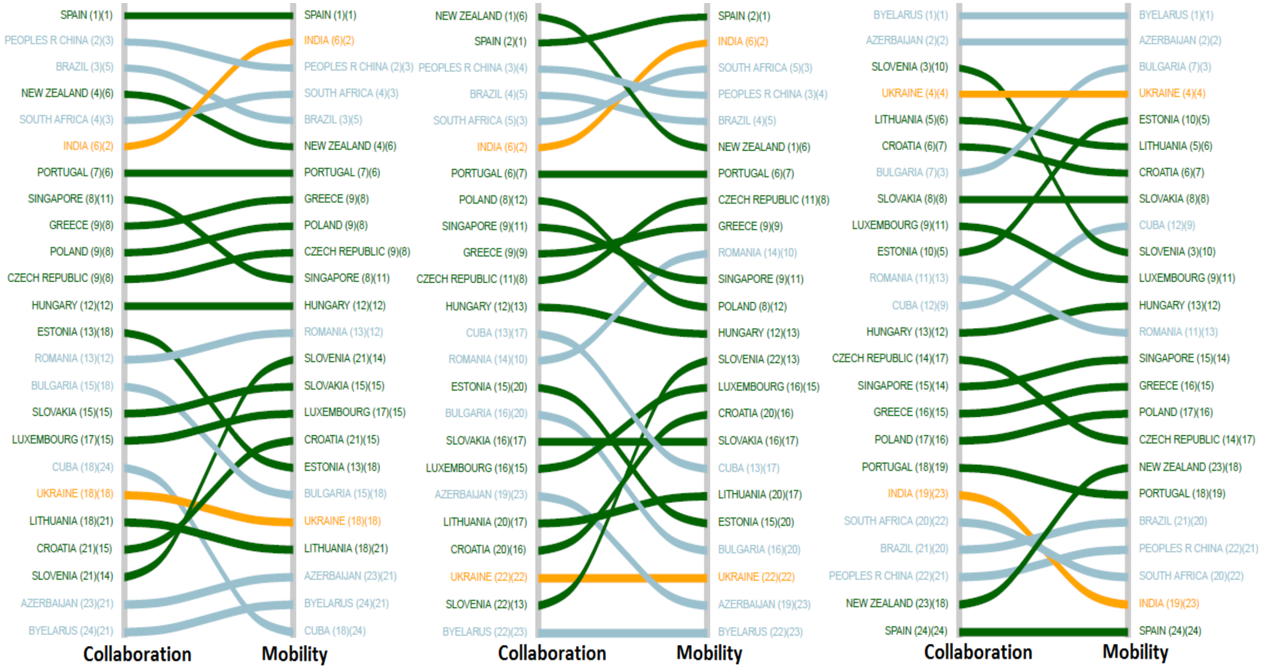

Figure 2. Closeness, betweenness and clustering coefficient for scientifically proficient countries

The positions of scientifically proficient countries fluctuate more than those of advanced countries (Figure 2). India ranks higher in closeness and betweenness in the mobility network than in the collaboration network, but their group of partners in collaboration is more cohesive than in mobility. That means that the country acts as a key bridge in the mobility network, but its mobility neighbors do not form tightly group around it. A similar situation appears in Slovenia and Croatia, whereas Estonia, Bulgaria, and Cuba have roles that are more important in international collaboration than in mobility.

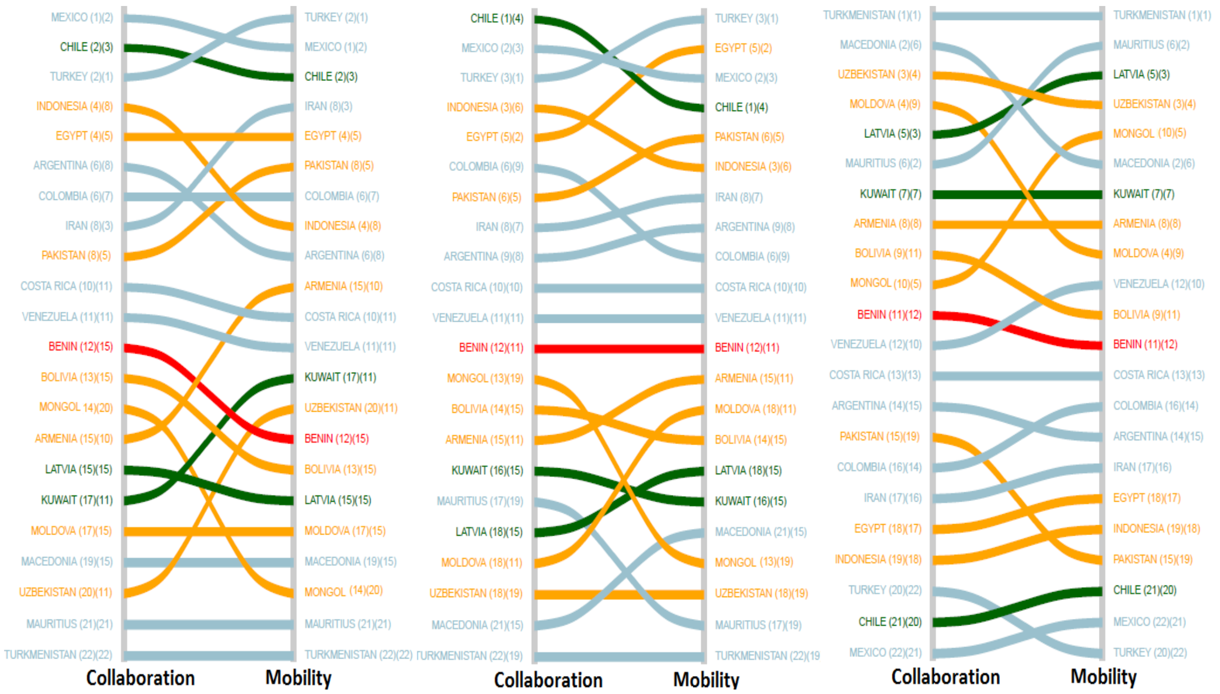

Figure 3. Closeness, betweenness and clustering coefficient for scientifically developing countries

In the group of developing countries (Figure 3), Mexico and Chile rank highly in closeness and betweenness, but their slight drop between collaboration and mobility demonstrates that they can build and facilitate more connections in collaboration than in mobility. Turkey shows the opposite pattern, being more important in mobility. These three countries show the lowest cohesion with their partners in comparison with the rest of the group. Iran's, Pakistan's, and Armenia's closeness and betweenness rank in mobility suggests than they have most important roles in linking researchers through mobility than collaboration. A striking case is Turkmenistan, which ranks at a top position by cluster coefficient and in the lowest position in closeness and betweenness, suggesting that while it does not act as a bridge in the network, it does forms a tight collaboration relationship with its partner compared to other countries in its group. Overall, developing countries are less stable between networks that advanced and proficient countries.

Countries classified as lagging are the most numerous (80 countries) and for visualization, we only show the top 30 in each indicator, explaining the lack of correspondence among the appearance of countries (Figure 4). The only high-income country appearing in this sample (among the six marked as lagging) is Saudi Arabia, which has an important role in closeness and betweenness, especially in the mobility network, whereas Malaysia leads in these two indicators in both networks. However, neither Malaysia nor Saudi Arabia establishes a tight relationship with their partners in comparison with the rest of lagging countries. In general, many countries' position changes dramatically, indicating that their positions and roles differ between the networks.

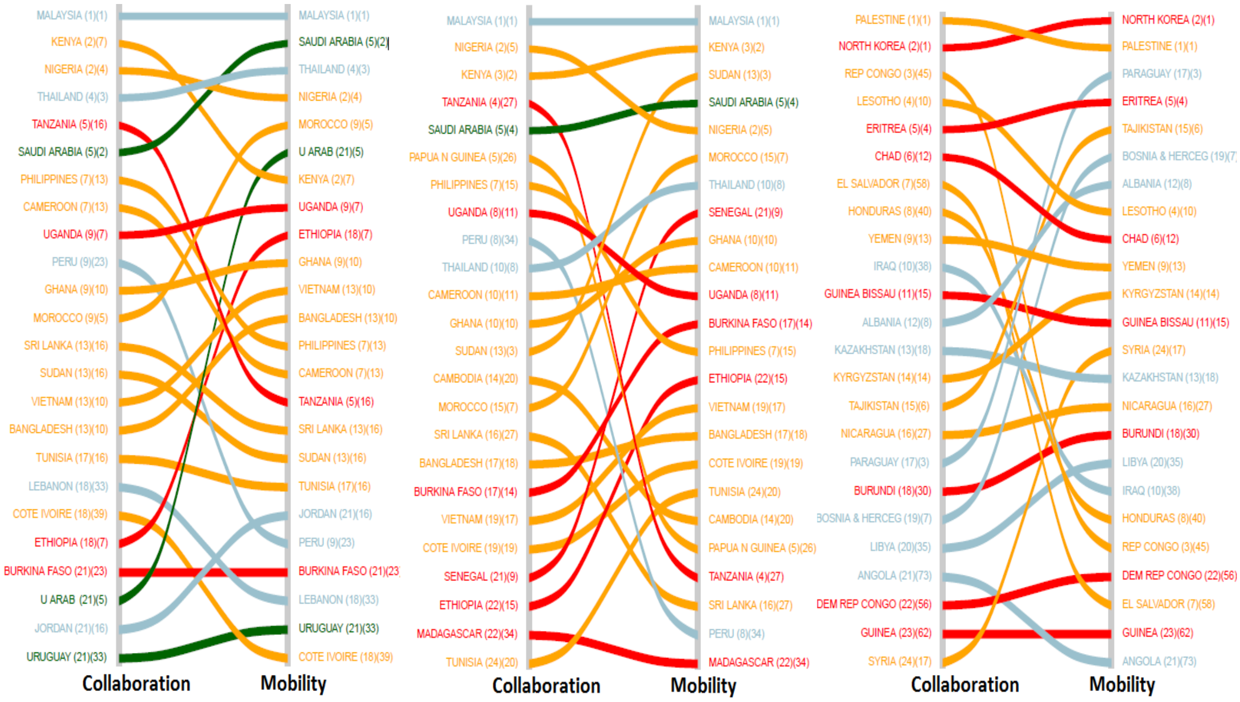

Figure 4. Closeness, betweenness and clustering coefficient for scientifically lagging countries

Both the collaboration and mobility networks demonstrate the importance of geographical proximity in the topology of the network (Figure 6). Historical and political linkages are also apparent—for example, the links between countries and colonies as well as linkages in former Soviet countries. Language also matters, as demonstrated by the connections between Spain and Latin American countries. The two networks the most advanced countries occupy the center, but the change in the position of some Asiatic countries such as India, Taiwan, Singapore, Malaysia or the position of South Africa, Nigeria, Kenya in the two networks reveal significant differences in partnership preferences but also in policies that promote the intensification and expansion of international scientific networks.

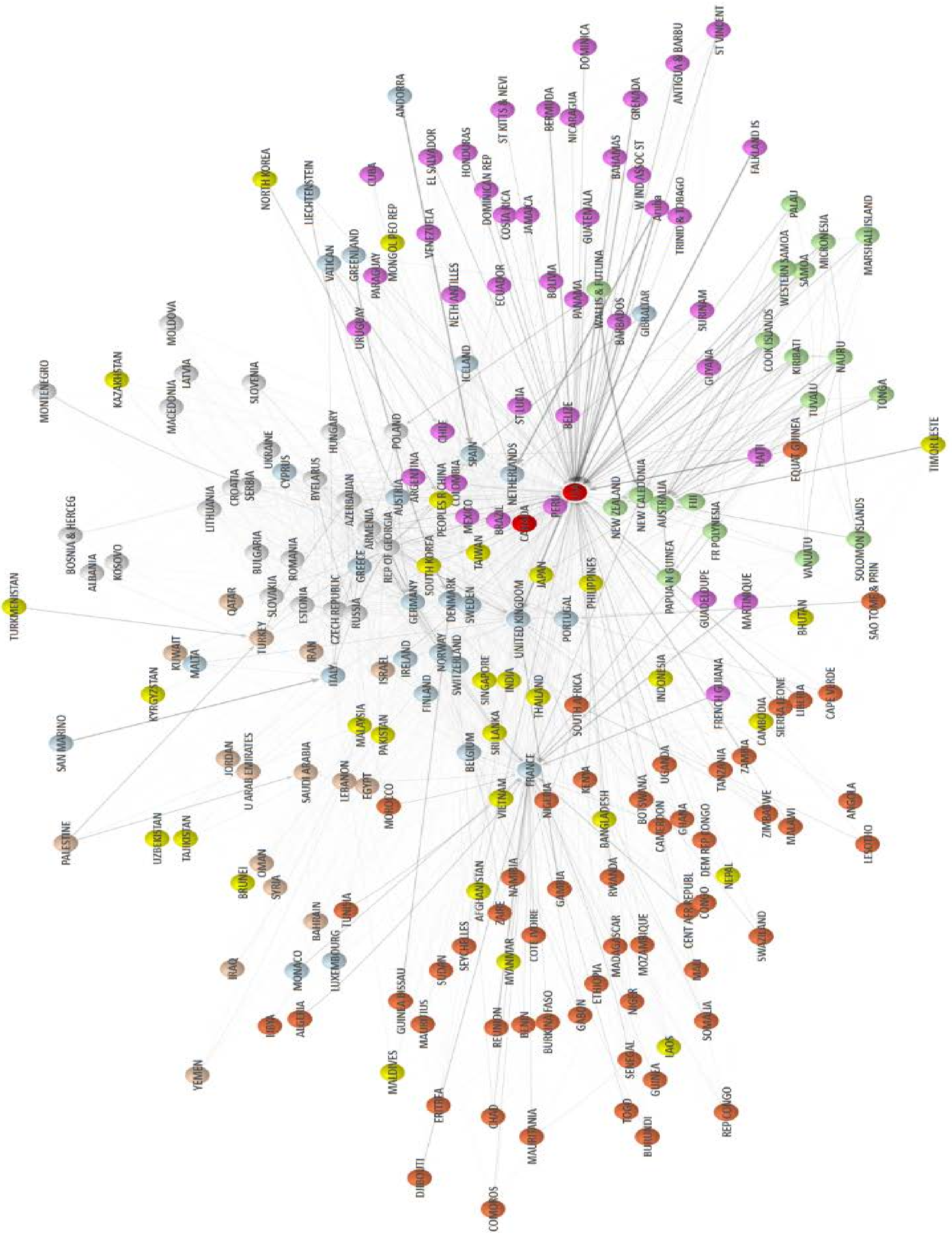

Figure 5. Network of international collaboration

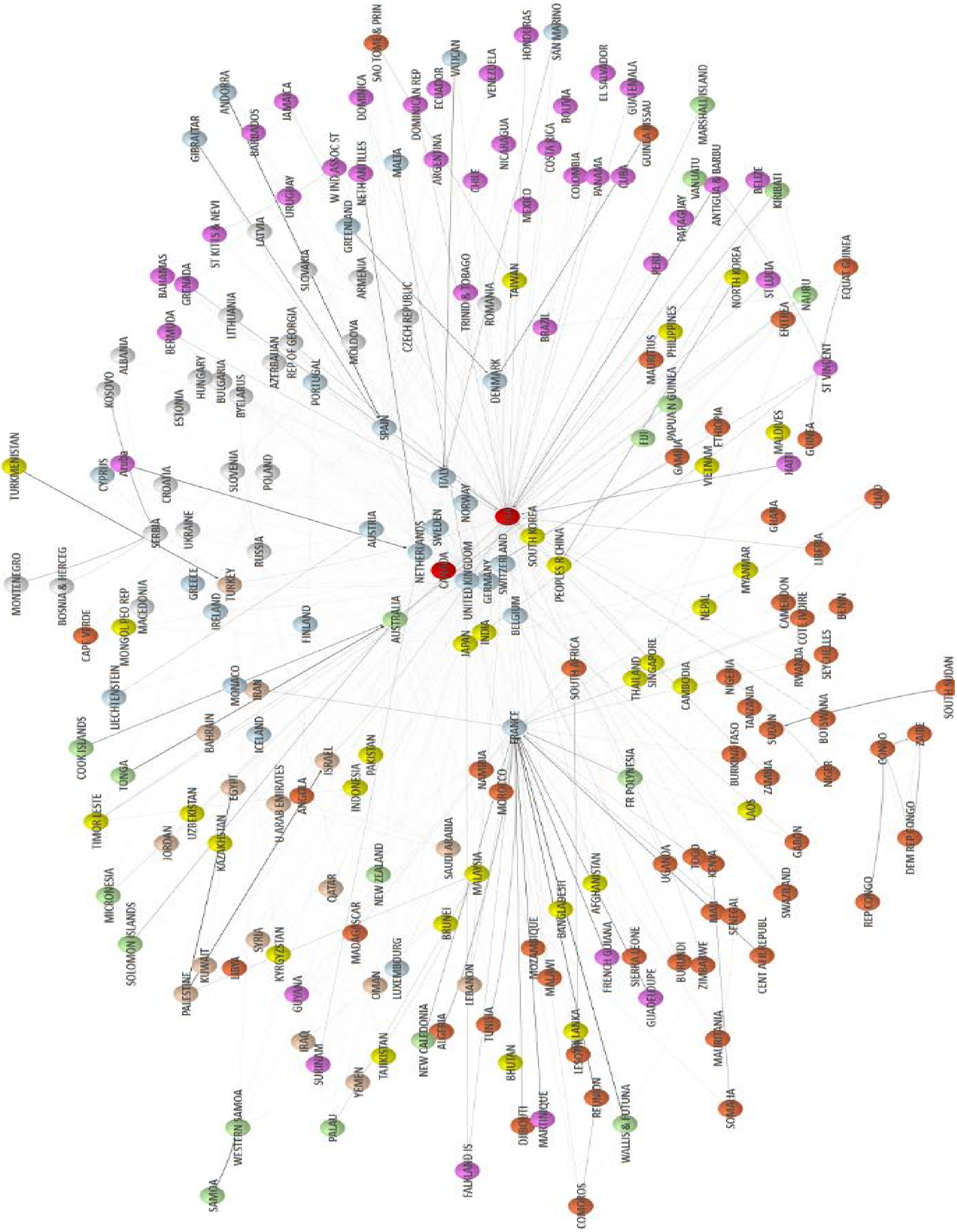

Figure 6. Network of mobility

Discussion and Conclusion

This study presents a preliminary comparison of the diversity and complexity in establishing international relationships in terms of collaborative papers and the flow of mobile researchers. The results show that researchers establish more international relationships through collaboration than mobility. To a certain extent, networks show a high level of local clustering and a small average number of steps between actors, which fits with the model of the small world defined by Watts and Strogatz (1998) and the results of Wagner (2005).

However, the mobility network forms tighter structures with fewer links than the collaboration network, making this a distinctive feature between the two networks, despite the diameter of collaboration network is tighter than in mobility network. A country's position in collaboration and mobility is not always consistent in both networks, revealing that positions and roles of countries in terms of capacities to be accessible to other nodes (closeness), in transferring knowledge (betweenness), or to be cohesive with their neighbours (clustering) may differ. Where stability between the networks exists, it is mostly associated with high-income countries, while upper-middle, lower middle, and low-income countries are more likely to change positions. Unlike countries in the scientifically advanced group that function as hub nodes, the role of lagging countries is unstable in both networks. However, the stability of a small subset of countries as bridge nodes, such as India, Malaysia, Turkey and South Korea, builds stronger connections in the mobility network. We also confirm the important role of Kenya and Nigeria, as noted in previous studies relating to collaboration networks (Adams et al., 2014). Although leading research economies tend to attract more researchers in terms of collaborative papers and mobility, the results suggest that there are complex patterns of knowledge circulation. The representation of the two networks also displays interesting drivers that reveal affinities between different economies based on linguistic, historical, as well as political and cultural linkages.

Further analysis is needed to overcome some of the limitations of this study and respond to other important questions related to the capacities and influences of countries in networking science. For example, the analysis of co-affiliations in Web of Science has an inflationary effect on the traditional measurements of collaboration based on affiliations creating some overlaps when we are comparing collaboration and mobility. There is a major general limitation of collaboration analysis based on author affiliations and further analysis should be done to minimize this effect (Hottenrott and Lawson, 2017). We intend to complement our analysis with a time component, allowing us to analyze some key points: authors' choices regarding institution address selection from publication data and the evolving relationships for exploring how these more fine-grained temporal networks compare to the entire network presented here. We will explore different approaches to community detection in networks to identify the core and groups of vertices having the highest probability of holding a great deal of influence over the organization of the periphery of the network. We plan to analyze not just mobility, but also leadership, including changes in the positions that authors occupy in the bylines of co-authorship, their impact, the institutional reputation of destinations, and the capacity to develop or reinforce thematic research into their institutions and countries. In addition, the analysis of other factors that would influence the mobility and collaboration of researchers such as cultural, linguistic and geographical proximities will be analyzed. Overall, this analysis presents an introduction to what is possible through comparison of international collaboration and mobility networks, and serves as a foundation for future analysis.

Acknowledgments

Financial support from Mobility Program 'Salvador de Madariaga 2016' and State Programme of Research, Development and Innovation oriented to the Challenges of the Society (Ref. CSO2014-57770-R) funded by the Ministry of Economy and Competitiveness of Spain and the Science of Science Innovation and Policy program of the National Science Foundation in the United States (NSF #1561299).

References

Adams, J., Gurney, K., Hook, D., & Leydesdorff, L. (2014). International collaboration clusters in Africa. *Scientometrics* 98 (1), 547-556.

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

Batagelj, V. and Mrvar, A. (1997), "Program package Pajek/Pajek-XXL", available at: http://mrvar.fdv.uni-lj.si/pajek/ (accessed December 1, 2016).

Beaver, D.D. (2001). Reflections on Scientific Collaboration (and its study): Past, Present and Future. *Scientometrics* 52 (3), 365-377.

Caron, E., & Van Eck, N.J. (2014). Large scale author name disambiguation using rule-based scoring and clustering. In Proceedings of the *19th International Conference on Science and Technology Indicators* (pp. 79-86).

Chinchilla-Rodríguez, Z., Benavent-Pérez, M., Miguel, S., Moya-Anegón, F. (2012). International Collaboration in Medical Research in Latin America and the Caribbean (2003-2007). *Journal of the American Society for Information Science and Technology*, 63 (11), 2223-2238

De Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory social network analysis with Pajek*. Cambridge University Press.

Freeman, L.C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1), 35-41

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3), 215–239.

Glanzel, W. (2001) National characteristics in international scientific co-authorship relations. *Scientometrics* 51(1), 69–115

Gazni, A., Sugimoto, C.R., Didegah, F. (2012). Mapping World Scientific Collaboration: Authors, Institutions, and Countries. *Journal of the American Society for Information Science and Technology*, 63(2), 323–335

Gibson J, McKenzie D (2012) The economic consequences of "Brain Drain" of the best and brightest: microeconomic evidence from five countries. *The Economic Journal* 122 (560), 339–375.

Hanneman, R.A. & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of Riverside.

Hottenrott, H., & Lawson, C. (2017). A first look at multiple institutional affiliations: a study of authors in Germany, Japan and the UK. *Scientometrics*, 111 (1), 285-295

Kamada, T. & Kawai, S. (1989). An algorithm for drawing general undirected graphs. Information Processing Letters, 31(1), 7-15.

Moed, H. & Halevi, G. (2014). A bibliometric approach to tracking international scientific migration. *Scientometrics*, 101(3), 1987-2001.

Newman, M. E. J. (2002). Assortative mixing in networks. *Physical Review Letters* 89 (20) 208701

OECD. (2008). *The Global Competition for Talent: Mobility of the highly skilled*. OECD Publications.

OECD. (2010*), Measuring Innovation: A New Perspective*, OECD Publishing, Paris.
Scellato, G., Franzoni, C., and Stephan, P. (2015). Migrant's scientists and international networks. *Research Policy*, 44(1), 108-120
Sonnenwald, D.H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology* 41, 643–681
Sugimoto, C.R., Robinson-Garcia, N. & Costas, R. (2016). Towards a global scientific brain: Indicators of researcher mobility using co-affiliation data. In *OECD Blue Sky III Forum on Science and Innovation Indicators*, Ghent, September 19-21. http://arxiv.org/abs/1609.06499
Wagner, C. S., Brahmakulam, I., Jackson, B., Wong, A., & Yoda, T. (2001). *Science and Technology Collaboration: Building Capacities in Developing Countries*. Santa Monica, CA: RAND.
Wagner, C. & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, 34(10), 1608-1618
Wasserman, S. & Faust, K. (1999). *Social network analysis-methods and applications*. Cambridge: Cambridge University Press.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684), 440–442.
World Bank. World Development Indicators. Retrieved from: http://wdi.worldbank.org/tables/ on Oct 26, 2016.
Zitt, M., Bassecoulard, E. & Okubo, Y. (2000). Shadows of past in international cooperation: Collaboration profiles of the top five producers of science. *Scientometrics, 47*(3), 627-657.