INTERNATIONAL COLLABORATION IN SCIENCE:
THE GLOBAL MAP AND THE NETWORK

Loet Leydesdorff, Caroline S. Wagner, Han-Woo Park and Jonathan Adams

Loet Leydesdorff (Ph.D. Sociology, M.A. Philosophy, and M.Sc. Biochemistry) is professor at the Amsterdam School of Communication Research (ASCoR) of the University of Amsterdam. He is visiting professor of the Institute of Scientific and Technical Information of China (ISTIC) in Beijing and honorary fellow of the Science and Technology Policy Research Unit (SPRU) of the University of Sussex. He has published extensively in systems theory, social network analysis, scientometrics, and the sociology of innovation (see for a list of publications at http://www.leydesdorff.net/list.htm).
http://orcid.org/0000-0002-7835-3098

Amsterdam School of Communication Research (ASCoR)
Kloveniersburgwal 48, 1012 CX Amsterdam, The Netherlands
http://www.leydesdorff.net
loet@leydesdorff.net

Caroline S. Wagner, an expert in the field of science and technology and its association to policy, society, and innovation, holds the Ambassador Milton A. and Roslyn Z. Wolf Chair in International Affairs at the John Glenn School of Public Affairs, The Ohio State University, Columbus, Ohio. Dr. Wagner earned her doctorate from the University of Amsterdam in Science and Technology Dynamics with a focus on collaborative research networks; she received her Master of Arts degree in Science, Technology and Public Policy from George Washington University; and she holds a Bachelor of Arts from Trinity College, Washington, DC.
http://orcid.org/0000-0002-1724-8489

Amb. Milton A. & Roslyn Z. Wolf, International Affairs
John Glenn School of Public Affairs, The Ohio State University
1810 College Av. N, Columbus, Ohio 43210, USA
wagner.911@osu.edu

Han-Woo Park is working as an associate professor in the Department of Media & Communication, Yeungnam University, South Korea. Over the past several years, he has contributed important works in webometrics from the perspective of social-network analysis. He was formerly principal investigator for an international research project entitled ‘Investigating internet-based politics using e-research tools’ (April 2009 to August 2011), funded by the Korean government’s WCU (World Class University) Program. He is a director of Cyber Emotions Research Institute and President of the Asia Triple Helix Society.
http://orcid.org/0000-0002-1378-2473

Department of Media & Communication, Yeungnam University
280 Daehak-Ro, Gyeongsan, Gyeongbuk 712-749, South Korea
hanpark@ynu.ac.kr

Jonathan Adams was the founder of Evidence Ltd, and is now director of Research & Development for Thomson Reuters. He has worked at King’s College London (1979-1980), University of Newcastle upon Tyne (1980-1983), University of Leeds (1983-1989) and Imperial College London (1989-1992). He was a member of the science policy staff of the UK Advisory Board for the Research Councils from 1989-1992. Evidence carries out research evaluation in Europe, Australia, Brazil, Canada, China, India, Malaysia, Saudi Arabia, Singapore and the USA.
http://orcid.org/0000-0003-3893-8128

Evidence Ltd, Thomson Reuters, UK
http://researchanalytics.thomsonreuters.com/grr
http://science.thomsonreuters.com/evidence
jonathan.adams@thomsonreuters.com

Nota: Este artículo puede leerse traducido al español en:

Article received on 10-01-2013
Approved on 31-01-2013
Abstract
The network of international co-authorship relations has been dominated by certain European nations and the USA, but this network is rapidly expanding at the global level. Between 40 and 50 countries appear in the center of the international network in 2011, and almost all (201) nations are nowadays involved in international collaboration. In this brief communication, we present both a global map with the functionality of a Google Map (zooming, etc.) and network maps based on normalized relations. These maps reveal complementary aspects of the network. International collaboration in the generation of knowledge claims (that is, the context of discovery) changes the structural layering of the sciences. Previously, validation was at the global level and discovery more dependent on local contexts. This changing relationship between the geographical and intellectual dimensions of the sciences also has implications for national science policies.

Keywords

Register for free at https://www.scipedia.com to download the version without the watermark of both China and Japan have gained a synergy from foreign co-authorship relationships. But it is still debatable whether international collaboration is positively associated with the quality of scientific outputs in terms of citation impact when controlling for countries and fields (Persson et al., 2004; Persson, 2010).

Coauthorship relations are a most formal indicator of international collaboration. Scientific collaborations may lead to a number of outcomes of which a co-authored paper is only one (Laudel, 2002; Katz; Martin, 1997). However, from the perspective of the development of the sciences as publication systems, the submission of manuscripts containing new knowledge claims is the crucial outcome. Furthermore, we acknowledge that coauthorship in itself does not imply that collaboration has occurred (Woolgar, 1976). It represents outcomes that the listed authors jointly view as notable, which serves as a socio-cognitive filter on the multitude of relations in the social context of discovery (Melin; Persson, 1996).

No researcher unnecessarily shares authorship and thus collaborative publication can be considered as an indicator of esteem and shared intellectual contributions. From a methodological perspective, coauthorship counts have the advantage of being reproducible over time and traceable.
year-on-year. The network of coauthorship relations offers a perspective on the ranks and positions of countries which provides an alternative to ranking shares of publications and citations.

Wagner & Leydesdorff (2005) suggested that international collaboration tends to free scholars from local constraints such as funding by national government agencies and social (linguistic, cultural) contexts having a direct impact on intellectual agendas. Wagner (2008) hypothesized the emerging layer of international collaborations as a “new invisible college” (cf. Crane, 1972). Leydesdorff & Wagner (2008), however, noted the formation of a central group of highly functioning nations while other nations tend to remain peripheral, possibly reinforcing a core-periphery model originally proposed by Ben-David (1971; cf. Choi, 2012; Schott, 1991). Using network statistics and cosine-normalization, these authors identified a core set of 12 European nations, the USA, and Russia in both 2005 and 2006, whereas other countries (e.g., Canada, China, and Portugal) could be considered at that time as peripheral. Language can also be associated with disadvantages in terms of access, particularly in the humanities and the social sciences (Larivière et al., 2006), since most bibliographic databases are focused on English literature.

In this study we present an update of the network for 2011 using the most recently available edition of the Science Citation Index (SCI). As previously, we used the DVD version of this index containing 3,744 journals. This selection from the 8,336 journals covered by the Science Citation Index–Expanded (SCI-E) at the Web of Science (WoS), can be considered as the most policy-relevant because it includes the most elite and highly cited of the refereed journals. The same data is, for example, used for the Science and Engineering Indicators series of the National Science Board of the USA (NSB, 2012, at pp. 5-37 ff.), which also includes an index of international collaborations for 2010 in tabular format. Our study provides complementary network and visualization techniques that enable the user to envisage the effects of this globalization and additionally to zoom in to specific regions and/or networks of specific nations (Wagner et al., in preparation).

**Methods and materials**

One of us downloaded the entire set of the DVD-version of the Science Citation Index 2011; this data was then brought under the control of relational database management (in the dbf-format using Flagship v7). The data contains 1,042,654 papers of which 778,988 fulfill two conditions: (i) a country address is provided and (ii) they are part of the subsets of (719,327; 69.0%) articles, (37,685; 3.6%) reviews, and (29,989; 2.9%) letters. Ephemera (such as editorial materials and meeting abstracts) were not included in our analysis. In the download, 254 country names could be distinguished, of which 201 valid entities were used as variables to the (778,988) documents as units of analysis. More than 99% of this data is in English!

An asymmetrical matrix of documents versus countries was saved as a systems file in SPSS (v20) for generating, among other things, a cosine matrix between the 201 variables (countries). UCINet (v6.28) was used to generate an asymmetrical co-authorship matrix among countries (after changing all values to binary) where a record with three addresses in

**Figure 1.** Map of international collaborations; the size of each node is proportional to the logarithm of the number of fractionally counted papers. Only countries with more than 500 papers are included. The descriptors of the nodes contain the number of fractionally counted papers. Available at http://www.leydesdorff.net/intcoll/intcoll.htm
country A and two addresses in country B is counted as a single relation between these two countries. (An affiliations routine in social-network analysis would otherwise count this as $3 \times 2 = 6$ relations.) Additionally, the papers were fractionally counted: fractional counting means attribution of each address to a paper proportional to the number of addresses provided in the byline of the article. For example, if two of the three addresses are in country A, the paper is attributed for $2/3$rd to this country and for $1/3$rd to country B.

Among these papers 193,216 (that is, 24.8% of the 778,988 documents under study) were internationally coauthored with 825,664 addresses (39.3% of 2,101,384). Note that these numbers are somewhat greater than but not substantially different from 2005, with 23.3% of the papers internationally coauthored carrying 36.5% of the addresses (Wagner; Leydesdorff, 2008, p. 319).

Both the co-occurrence matrix and the cosine-normalized matrix were further processed in Pajek² and VOSViewer³ for the network analysis and visualization, respectively. Using the GPS Visualizer at http://www.gpsvisualizer.com/map_input?form=data and thresholds of minimally 500 fractionally-counted papers for each country and 500 international relations for each link, a global map of international collaborations was generated; this map is available at http://www.leydesdorff.net/intcoll/intcoll.htm. The links were not weighted according to the number of coauthorship relations because this would overload the visual. Instead, a legend is inserted, and in the interactive format one can click on each link to obtain the number of collaborations in a descriptor of the Google Map.

The global map of science at http://www.leydesdorff.net/intcoll/intcoll.htm provides users with an overview and all the functionalities of a Google Map, such as zooming and tagging. For example, one can click on each node and obtain the number of internationally coauthored papers based on fractionally counted papers in the set of 778,988. The links are all counted as unity (as explained above). Links can also be clicked or read from the legend. The nodes are sized proportionally to the logarithm of the number of papers.

As figure 1 shows, 440 of the 12,339 links between nations surpass the threshold of more than 500 co-authorship relations (of the $[201 \times 200 / 2] = 20,100$ possible links); 53 nations are involved. Thus, international collaboration is heavily concentrated. As an example, the link between Canada-Sweden is highlighted in the descriptor and centered in the legend table to figure 1. Visual inspection of the map shows that from the sub-Saharan countries only South Africa contributes, and within Latin America participation is limited to Brazil, Argentina, Chile, Venezuela, and Mexico (Wagner; Wong, 2012).

The network among EU nations is very dense. Integration makes the USA appear to operate as another member state of the EU. (One can zoom in using Google Maps online.) However, China has now become the first partner of the USA in terms of international co-authorship (that is, 12,950 integer-counted papers against 11,337 coauthored with an address in the UK). Recent accession countries (e.g., Poland and Bulgaria) are not connected because of Malta and Luxembourg. The geographical map shows that from the sub-Saharan countries only South Africa contributes, and within Latin America participation is limited to Brazil, Argentina, Chile, Venezuela, and Mexico. (Wagner; Wong, 2012).

The network among EU nations is very dense. Integration may share a common pattern of relations may not relate intensively, but they are all counted as unity (as explained above). Links can also be clicked or read from the legend. The nodes are sized proportionally to the logarithm of the number of papers.
Figure 3. The strong component of 42 nations in the center of the network (no normalization implied). Nodes are normalized in terms of their numbers of relations (i.e., degree distributions); VOSViewer used for clustering, coloring, and mapping. Available at http://www.vosviewer.com/vosviewer.php?map=http://www.leydesdorff.net/intcoll/core42map.txt&network=http://www.leydesdorff.net/intcoll/core42net.txt&n_lines=3000&label_size=1.35

normalization for size captures this comparison among distributions because the cosine can also be considered as a proximity measure (independent of the mean size of distribution, but without the reference to the mean; cf. Ahlgren et al., 2003).

Figure 2 shows the network of international coauthorship relations among 190 countries. Some smaller nations (such as Kosovo, Gibraltar, and the Netherlands Antilles) were removed because they tend to distort the figure by pulling the center towards outliers. The map shows the Anglo-American countries on the right side of the figure as similar in their collaboration patterns. In this projection, the Asian nations are positioned towards the bottom-left side — with the exception of Japan — with the nations of the Middle East as a nearby cluster.

Continental Europe is in the middle. The European position is caused by the dense network of collaborations among the core EU nations (such as France, the Benelux countries, and Germany). Portfolios of EU nations are influenced by the funding of the European Commission’s science and cohesion policies requiring collaboration. Japan is not visible on this map because its node is hidden behind France in the center area; the node and label for Japan can be made visible by choosing the (alternative) “label view” in VOSViewer. Certain other nations such as Argentina, Brazil, and Mexico are also related to this set, whereas Chile, for example, is more exclusively related to Spain. The somewhat specific positions of Italy and Austria at the peripheries of this map are noteworthy showing that the accession countries of Central and Eastern Europe are integrated in a triangle involving these two nations and Germany.

c. Center and periphery in the network

Figure 3 shows the network among 42 nations forming a strong component in the network of international coauthorship relations in 2011.

This figure shows the major players in the network in terms of international coauthorship relations. In contrast to the ranking of shares of publications in terms of addresses — China is also second behind the USA in terms of fractional counts — this figure shows, among other things, that China is not (yet) so active in terms of international coauthorship as are advanced industrial countries (e.g., the UK and Germany; National Science Board, 2012, at p. 5-37; cf. Plume, 2011). However, in contrast to data examined in 2005/2006, China is now part of the central group.

The polar position of France (at the bottom left) is noteworthy and can be considered as a consequence of its leading position (along with Spain) in collaborations with Mediterranean and Romance-language-speaking countries. Despite the nearly global use of English as the language of research publication (99.1% in this data), there are still distinct collaborative groupings of Francophone countries in Africa (Adams; King; Hook, 2010; Adams et al., in preparation)
and Luso-/Hispanophone nations in central and South America. These networks point to cultural and economic factors underlying regional differentiation in the global patterns.

d. The international environments of nations

As noted, individual nations may not be visible on the global map at http://www.leydesdorff.net/intcoll/intcoll.htm because of insufficient representation with regard to thresholds. Regional analyses, with more relaxed thresholds on volume of activity and collaboration, enable the user to extend this analysis and show how countries may become local hubs to emerging regional networks (Adams; King; Hook, 2010; Adams et al., 2011).

Indonesia, for example, has 559 papers in the set, but fractionally counted these add up to only 227.9 coauthored documents. Using Pajek (or any other network analysis program), the user can bring the co-authorship neighborhood of a specific nation to the fore, as in figure 4 for Indonesia: 86 countries are included in this so-called ego-network, but with (sometimes single) co-authorship relations.4

Although not a major player in the global science system, Indonesia is strongly networked to the extent that on average each paper is coauthored 2.5 times (≈ 559/227.9). The main international relations are with advanced industrial neighbors in the Asian-Pacific region, the USA, and specific European nations. Many of these relations may be a consequence of scholars having studied abroad as postdocs or Ph.D. students.

Given the origin and readership of this journal, we were asked to pay additional attention to Latin America, Spain, and Portugal. Figure 5 provides the collaboration network among these nations including some which can be considered francophone (e.g., Haiti) or anglophone (e.g., Trinidad Tobago), but which one can expect to be integrated in the region.

Figure 5 first shows the much stronger connection between Spain and Portugal—as both EU member states— when compared with the linguistic relations overseas. Spain has remained a hub between the EU and Latin America more than Portugal (Glänzel et al., 2006). Relations among Chile, Brazil and Argentina are less developed than those between each of these countries and Spain (Presmanes; Zumelzu, 2003). Countries with languages other than Spanish or Portuguese are peripheral to this network as are some nations in central America. In summary, south-south collaboration remains peripheral when evaluated from the global perspective (Adams et al., in preparation).

Summary and conclusions

The network of coauthorship relations offers a perspective on the ranks and positions of countries which provides an alternative to ranking shares of publications and citations. The core group of collaborating nations is dominated by a subset of research-intensive Western European nations and the USA. This configuration was challenged during the 1990s and early 2000s by the arrival of new entrants at the global level. The analysis shows, all the nations of the world are now participating in this process of globalization. Whereas Leydesdorff & Wagner (2008) once feared that a small set of (approximately 14) nations could monopolize the network by reproducing historical patterns, the leading group has tripled to more than 40 nations (figure 3) in the last five years, suggesting a different dynamic operating at the global level. Thus, the development is more inclusive than before, with features more similar to an open system with some regional differentiation than the core-periphery grouping that characterized the global system in the past.

The globalization of co-authorship relations at current levels— with almost 25% of the relevant papers internationally coauthored, but carrying almost 40% of the institutional addresses in the
file— can be expected to have changed (or reflect changes in) the structure of science and the dynamics of knowledge creation in the core set. Whereas the context of discovery for generating knowledge claims was previously considered mainly a social context while the context of validation was envisioned at the global (or universal) level (Popper, [1935] 1959), nowadays the two contexts are increasingly intermingled.

Gibbons *et al.* (1994) hypothesized a third “context of application” that allows stakeholders to participate in the process of knowledge production and validation (cf. Lepori, 2011). National science policies based in institutions created in the 20th century may be less effective in influencing such a complex and adaptive system developing at the global level.

Notes

1. Addresses in England, Scotland, Wales, and Northern Ireland were recoded as “UK”.
2. 
Pajek is a network visualization and analysis program freely available for non-commercial usage at http://pajek.imfm.si/doku.php?id=download
3. VOSViewer is a program for network visualization freely available at http://www.vosviewer.com
4. The file for Indonesia is brought online for didactic purposes at http://www.leydesdorff.net/intcoll/indonesia.paj. The subsequent steps after opening the file in Pajek are as follows:
   1. Read the full network (“coocc201.net”; included in the file “indonesia.paj”).
   2. Network > Partition > k-neighbours; select node number and distance 1.
   3. Operations > Network + Partition > extract subnetwork 0-1; “0” for ego, “1” for k=1 neighbours.
   4. Partition > Make Cluster > 1 (only k=1 neighbours).
   5. Operations > Network + Partition > Transform > Remove Lines > Inside Cluster 1 (that is, links among k-neighbours).
   7. You may have to turn off labeling only the cluster under Options in the draw screen; otherwise one only sees the k-neighbours labeled.

Acknowledgments

Some authors acknowledge support from the SSK (Social Science Korea) Program funded by National Research Foundation of South Korea; NRF-2010-330-B00232

References
