Mapping Global Science Using International Co-authorships: A Comparison of 1990 and 2000

Caroline S. Wagner[1] and Loet Leydesdorff
Science & Technology Dynamics,
Amsterdam School of Communications Research (ASCoR),
University of Amsterdam

Abstract

Using the Science Citation Index (CD-Rom version) for 1990 and 2000, this paper analyses international co-authorships. The paper presents the methodology for identifying these co-authorships. Analysis is presented on observed linkages at the global level and on regional bases. The architecture of the network is further explored using statistical methods and factor analysis to reveal intense relationships as well as the core members of a global network. Findings show that, in the 10 years between 1990 and 2000, the global network has expanded to include more nations and it has become more interconnected. Regional networks show emerging hubs. Within the global network, a core set of countries has expanded from six in 1990 to eight in 2000.

1. Introduction

This paper analyses the dynamics of science as a global system by examining international co-authorships. It improves upon earlier analysis by including all the countries of the world, and by applying a range of tools, including social network analysis and factor analysis, to expose the network. We explore whether the network created by international links among scientists might be taking on the features of a global system overlaid upon and somewhat independent of the national systems.

There has been a rapidly growing literature discussing the increase in international linkages in science. Authors have approached the question from three perspectives: 1) scientometric analysis of the increase in the interconnectedness of scientists (examples include, Glänzel, 2001; Zitt, et al., 2000; Luukkonen, 1993; Okubo et al., 1992); 2) social sciences analysis of collaboration in general (Katz and Martin, 1997; Gibbons et al., 1994) and international linkages in particular (Schott, 1998; Stichweh, 1996); and 3) policy analysis of the implications of linkages for funding and outcomes (Advisory Council of Canada, 2001; Wagner et al., 2000).

The theories of why international collaboration is increasing appear to consider factors both internal and external to science. In addition, they address either the diffusion of scientific capacity or the interconnectedness of researchers, although, at times, the two features are not well delineated. In discussing reasons for the diffusion of scientific capacity, for example, Schott (1998, 1991), following Ben-David (1971) and Shils (1988), sees the progression related to a succession of countries that have acted as “centers” for world science, with countries at the periphery (often smaller countries) trying to emulate the organisation and orientation of scientific work at the center. As they emulate and adapt the practices of the core country, the capacity of the periphery countries grows.

A greater volume of literature addresses the increasing interconnectedness of researchers. Reasons offered for this phenomenon (ones highlighting factors internal to science) include Stichweh’s (1996) assertion, following Price (1963) that collaboration arises from the dynamics of internal differentiation of science into specialized disciplines. Galison has suggested that, at least for some sciences, the scale of investment is so large that no single nation will undertake it alone. These field-specific characteristics make some collaborations unavoidable. (Galison, 1987, quoted in Kim, 2002) Factors external to science offered as explanations for collaboration include geographical proximity and historical determinants explored by Zitt et al. (2000), and the ubiquitous mention of the rise of information and communications technologies as influencing the interconnectedness of everyone, not just those in the scientific community. (Gibbons et al., 1994 is just one example)

We are interested in examining both the diffusion of capacity and the interconnectedness of researchers at the network level since we view these as related phenomena. By looking at 1990 and

[1] Kloveniersburgwal 48, 1012 DX Amsterdam, Netherlands, wagner@pscw.uva.nl

Caroline S. Wagner and Loet Leydesdorff, *Mapping global science using international co-authorships: a comparison of 1990 and 2000*, In: Guohua Jiang, Ronald Rousseau, Yishan Wu (Eds), Proceedings of the 9th International Conference on Scientometrics and Informetrics – ISSI 2003, Dalian University of Technology, Dalian, China, 2003, 330-340.

2000, we sought to test the hypothesis that global science has grown more decentralized and more interconnected as well as more inclusive of new entrants.

2. Methodology

We began by examining the patterns of linkages among researchers from different nations. The Science Citation Index (SCI) produced by the Institute for Scientific Information (ISI) is the most reliable source for a comprehensive survey. We drew data from the Science Citation Index CD-Rom from 1990 and 2000.[2] The addresses are generally well standardized and can be compared over time. The addresses make it possible to study co-authorships using *countries* as the unit of investigation.[3] We did not distinguish among types of contributions (reviews, letters, proceedings, and journal articles) because we are seeking social connection to reveal the structure of the network regardless (at this point) of the scientific status of the output.

Counting and representing the patterns of linkages can be done in several ways. Absolute measures allowed us to see the overall size of the network, as well as which countries are in the center and which are at the periphery of the network. We used these measures to view the network at the global and regional levels. Relative measures normalize size and therefore provide insight into the strengths of links between countries. We also applied these measures in an effort to reveal the architecture of the network, and to illuminate the architecture of the relationship of countries. Multilateral measures take the global network into account based on the assumption that the expected numbers of linkages between countries occur in proportion to a country's share in the global network. We used to these to examine the core countries within the network.

We began by collecting into a single data set all papers produced for the relevant year and determining how to attribute papers to participating countries. Table 1 inventories this data set.

Table 1. Data used to create international network of co-authorships

Year	Unique documents in SCI	Addresses in the file	Authors for all records	Internationally co-authored records	Addresses, internationally co-authored records	Percent internationally co-authored documents
2000	778,446	1,432,401	3,060,436	121,432	398,503	15.6
1990	590,841	908,783	1,866,821	51,596	147,411	8.7

Since we are interested in using counts to see how the network is developing between countries, we used integer counting. During the counting phase, we applied a two-tiered analysis, first based on occurrences and then on co-occurrences. The data was placed into an asymmetrical matrix containing raw occurrence data that recorded all countries across one axis, and all articles on the other. A count is placed in the corresponding cell created by this matrix. The digit in the cell represents the number of unique addresses for each country appearing in the address line of the respective article. This occurrence table was used to conduct factor analysis in SPSS to identify patterns within the global network. We then converted the occurrence table into a binary matrix of only ones and zeros ("integer counting") in order to construct a symmetrical matrix of countries appearing on both axes, with the co-occurrence of addresses appearing in the corresponding cell. The co-occurrence table was also used to conduct social network analysis.

Once we collected all the data in the co-occurrence matrix for 1990 and 2000, we applied two types of analyses. Luukkonen et al. (1993) state that "The Jaccard measure underestimates the collaboration of smaller countries with larger ones, but the Salton measure underestimates the collaboration of smaller countries with each other." In a careful analysis, we find that Jaccard's measure does not serve us as well as the Salton Index. Jaccard's Index provides the intersection of the two countries as

[2] The CD-Rom version is preferable to the SCI Extended version found on the Internet. The Extended version is regularly updated, and therefore, it would be very difficult to repeat this analysis, even from day to day. Assuming that the extended index can change, the data set could be altered and therefore the results presented here could not be reproduced.

[3] The United Kingdom is considered here in its component parts. Addresses are provided as England, Scotland, Wales, and Northern Ireland, and each is handled as a separate political unit for the purposes of this analysis.

a percentage of the sum, while the Salton Index provides the intersection as a weighted percentage. But the difference is more than a factor two: whereas the Jaccard Index focuses on strong links in segments of the database (e.g., the strong relations between Croatia and Slovenia), the Salton Index organizes the relations geometrically so that they can be visualized as structural patterns of relations (Hamers et al., 1989 explores this in detail). Unlike the Pearson correlation, however, the Salton Index remains non-parametrical (Ahlgren et al., 2003; Leydesdorff & Zaal, 1988). Thus, in keeping with Glänzel (2001), we have used the Salton Index as a measure of the networked relatedness of countries.

Social network analysis allowed us to explore and visibly depict the intensity and dynamics of interrelationships among researchers from different countries at the multilateral level. Bringing the matrix of co-occurrences into Ucinet for analysis, and then into Pajek for the visualization of the networks, we sought to identify the clusters representing collaborations at a global and regional levels.

3. Findings: Comparison of Collaborative Networks

Within each of the three types of analysis applied, the data support our initial expectations: between 1990 and 2000, the global network has expanded (more players are involved), and it has become more interconnected (more links occur between players). The cluster created by scientifically advanced countries has expanded, but some other nations (e.g., the Arab countries in the Middle East) are grouped into otherwise disconnected networks. At the regional level, networks have expanded (more players), become more decentralized (increasing number of hubs), and are more interconnected (more linkages). Clustering retains features related to geographical proximity and historical relationships, but these are no longer the strongest features affecting links. Overall, the data supports the expectation that the science system is further developing into a network structure operating at the global level, one that is taking on dynamic characteristics.

At the global level, using network analysis to observe pairs of linkages, the network of interactions is very strong and highly interconnected. A large core of cooperating countries expands from 37 in 1990 to 54 countries in 2000. Regional networks show strong growth in linkages. Regions that had an active network in 1990 show the greatest growth by 2000, with the European Union and Accession Countries exhibiting a great deal of dynamism. Similarly, the Americas have a network which is heavily dominated by North America, but with smaller countries joining the network by 2000. In Asia, the most significant change is the movement of China from the periphery to the center of the Asian network. (Figures 1 and 2 are included as examples of the types of networks that we created.) Like the Americas, small countries that were not visible in 1990 have joined the Asian network. Africa shows emerging hubs in Kenya, Nigeria and Ethiopia; some smaller countries remain outside the regional network within Africa. The network in the region of the Middle East and North Africa shows very little growth over the decade.

Relational measures normalize the data and reveal the network architecture within the multilateral relationships. As discussed in the methodology section, we found the Salton Index to be the most illustrative of the overall hierarchy and structure at the network level, while the Jaccard Index focuses more on outlaying relations forming strong segments in the data. The Salton Index normalizes for the size of the participants while retaining a measure of the volume of the linkages between them. We took the co-occurrence matrix and applied the Salton Index equation to it to create a fully indexed set of data for 1990 and 2000. Then, we exported the data into Ucinet and into Pajek to develop visualizations of the hierarchies and structures within the global network.[4]

The results of the Salton Index (SI) analysis at the level of the core relationships among collaborative countries are revealing. By setting a threshold for relationships, it is possible to see that, even at a low threshold of SI $\geq$ 0.01, many countries fall out of the network. When the threshold is raised to 0.05, 42.1 percent of all countries fall out of the network in 1990; in 2000, the number is somewhat less at 35.4 percent. We are left with a component of 123 countries that are collaborating; 61 are bi-connected, that is, they are collaborating with more than a single other nation.

---

[4] For information about the Pajek software and its applications, see http://vlado.fmf.uni-lj.si/pub/networks/pajek/ [last visited, December 2002]

Some major changes during the 1990s can be noted. By 2000, (figure 3) the relations of the former colonial powers with their overseas territories have faded away with the exception of the relations within the British Commonwealth. The Francophone nations form the core of an emerging African cluster. South Africa is not part of this cluster: it belongs to the group of advanced nations. The more diffuse picture in 1990 (figure 4) can also be explained by a secondary network centered on the USSR, as a subset within the core set.

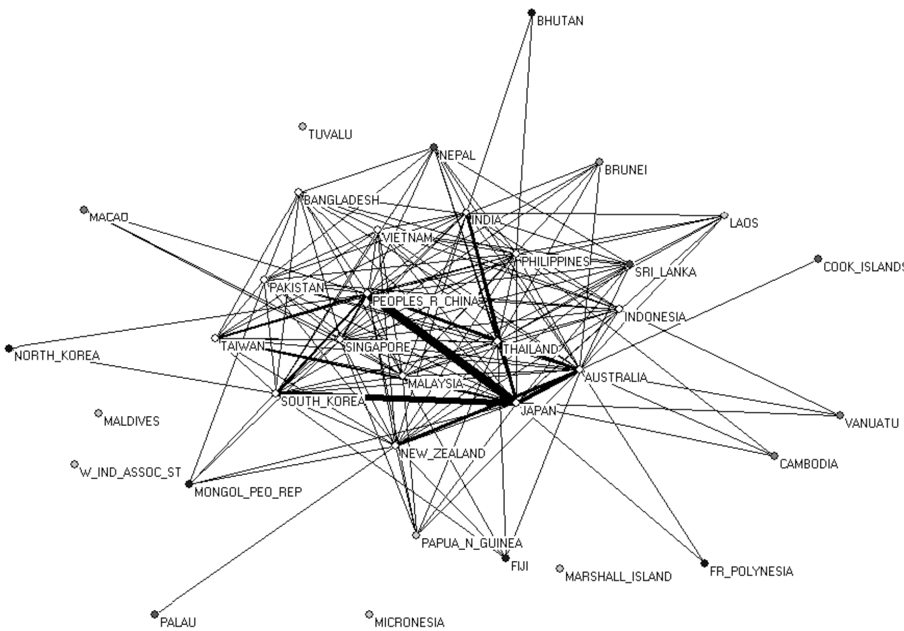

Figure 1. Network of Asia-Pacific country co-authorships, 2000

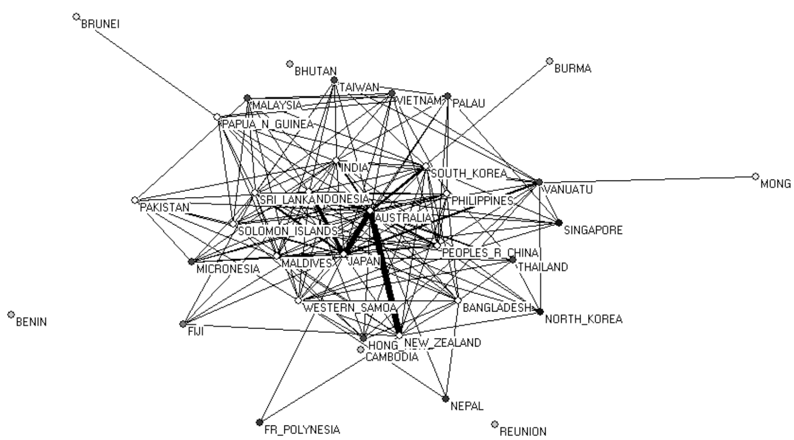

Figure 2. Network of Asia-Pacific country co-authorships, 1990

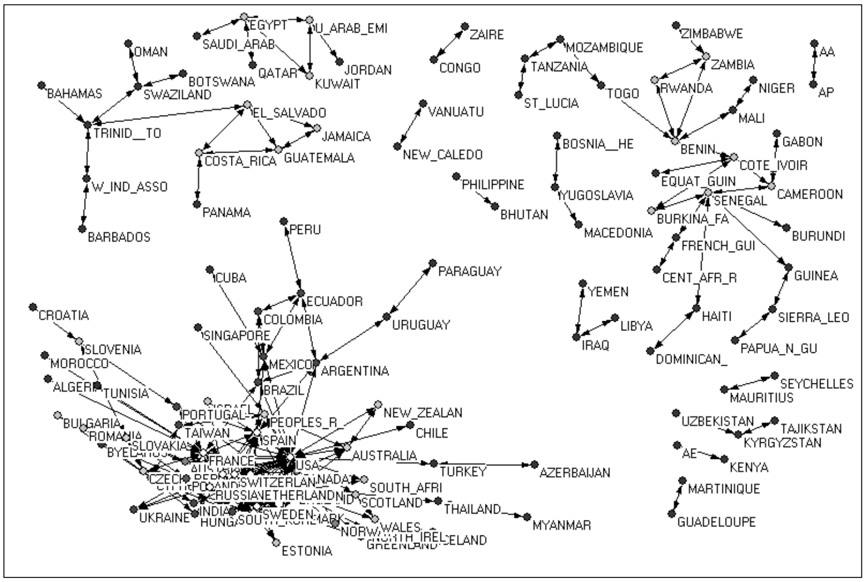

Figure 3. 128 countries collaborating in 2000 (Salton Index ≥ 0.05). (The cluster of eight core countries is indicated with white marks.)

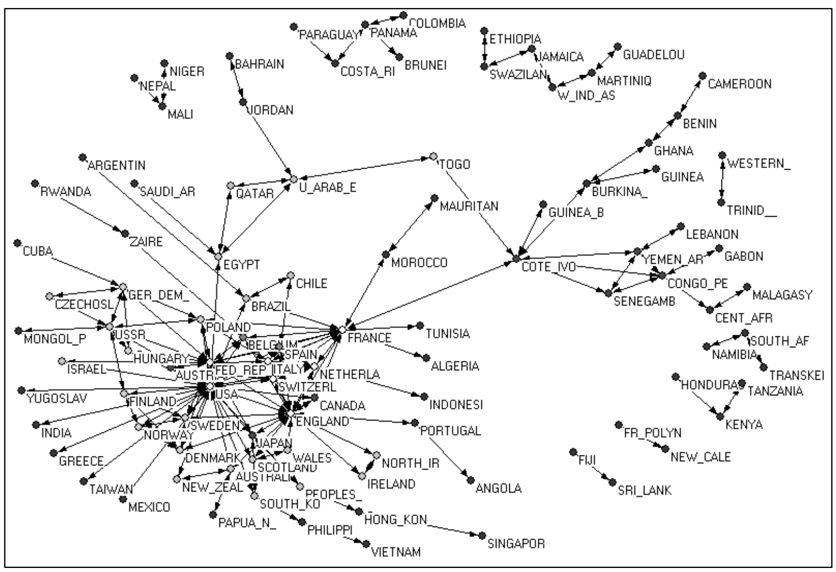

Figure 4. 103 countries collaborating in 1990 (Salton Index $\geq$ 0.05). (The cluster of six core countries is indicated with white marks.[5])

In 2000, the Eastern European part of the former Soviet group has now completely merged with the OECD set, particularly through Germany. This set of advanced nations also relates to a South American group, which is also more integrated than it was in 1990. The Caribbean and Central American networks, however, are separately organized in both years. The network of Arab countries has become more structured, but in 2000 this group is no longer related to the main grouping of more advanced countries.

---

[5] These visualizations are based on using the Fruchterman Reingold (1991) algorithm in Pajek (Batagelj and Mrvar, 2000).

We applied factor analysis to the asymmetrical matrix of co-authorships created using the SCI addresses from the 2000 and 1990 CD-Rom data. Factor analysis allows us to find a commonality of relationships among the variables in the data. We factor analysed this matrix forcing different numbers of factors to reveal different information about the structure and architecture of the relationships. In comparison to the observed data presented above, factor analysis enables us to recognize structural properties of correlation and variation that are not observable by inspection of the matrix level.

Factor analysis reveals that countries group in five distinct clusters that reveal geographic proximity or historical linkages. The breakdown of geographic groupings based on the solution for ten factors are:

1) a U.S. dominated cluster representing both proximity in the Pacific region and historical/political ties.

2) A Scandinavian cluster including Sweden, Denmark, Norway, Finland, and Estonia.

3) A continental European cluster that has 4 sub-clusters.

4) A second European cluster including France, Spain, and Belgium that relates to the Francophone and Latin world, including Senegal, Cameroon, Argentina, and Morocco.

5) A British Commonwealth cluster.

Figures 5 and 6 show the global network as a two factor structure for 2000 and 1990, respectively. On the above side of the picture in Figure 5, the Anglo-Saxon dimension is spanned between the U.S. and the various parts of the U.K. On the right side of the picture, one can observe the European dimension, gradually moving down from the U.K. to Germany and Russia. The third factor (not shown) opens up a second dimension within the European domain between France on one side and the Scandinavian countries on the other.

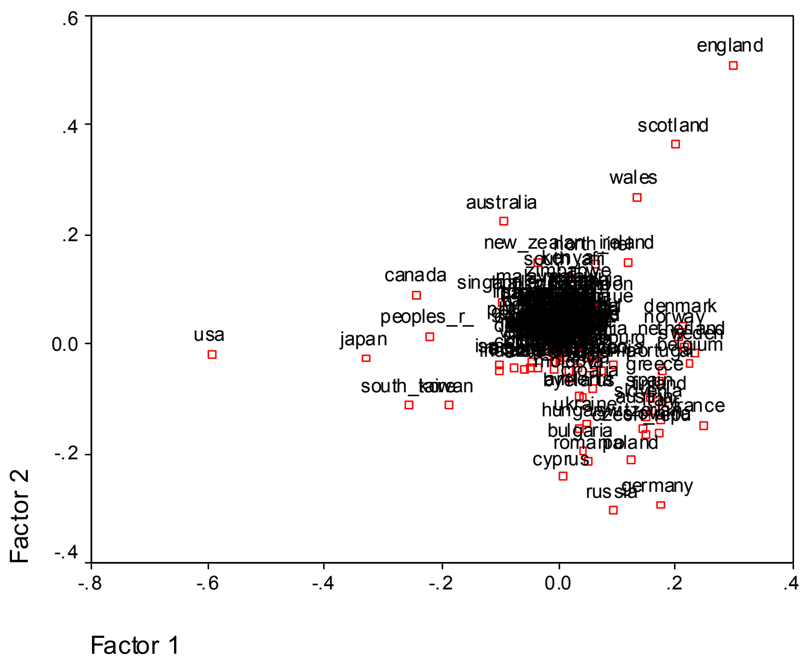

Figure 5. Factors One and Two used as dimensions for the mapping the rotated factor solution of the network of international co-authorship relations in 2000

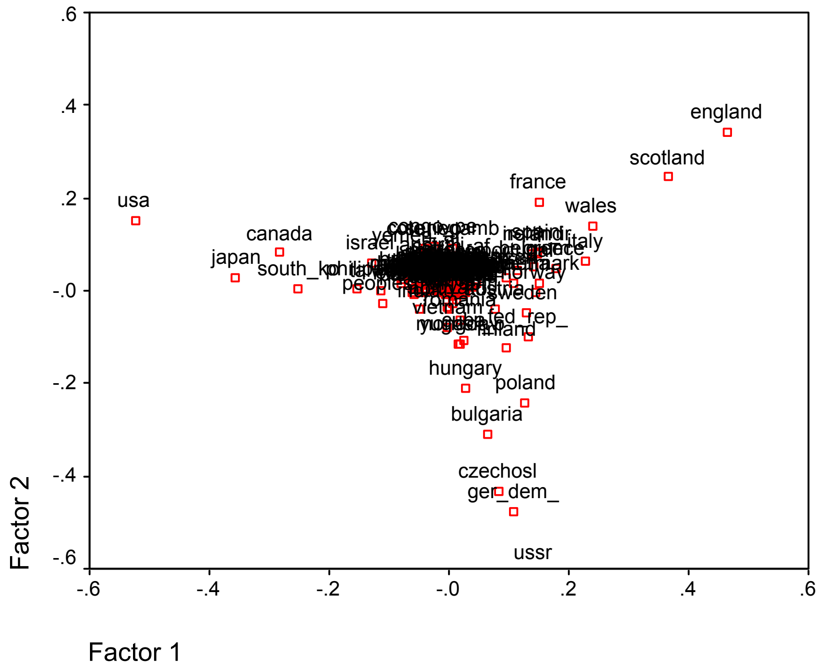

Figure 6. Factors One and Two used as dimensions for the mapping the rotated factor solution of the network of international co-authorship relations in 1990

In these representations, the peripheral countries occupy positions in the middle since individuals and institutions vary in their collaborations with the leading countries that span the network. This global net acts as a superstructure of international collaboration and competition to which the new entrants connect by creating linkages either directly or via intermediate hubs.

The disappearance of the Soviet Union has contributed to a further integration in the relevant networks of international co-authorship relations. The *position* of the Soviet Union in the international arena in 1990 was taken over by Germany in 2000. The European network, however, has become more complex. When using social network analysis for discovering a hierarchy (in terms of center and periphery) we found a strong cluster in terms of relations—including also the U.S. and Switzerland—but the factorial analysis shows that the core network is structurally differentiated and differentiating. The co-authorship relations with authors from Scandinavian countries, for example, have begun to play in the European theatre during this decade.

It is interesting to note that the U.K., the U.S., and Germany are counter-posed against one another in Figure 5 for 2000. This suggests that when researchers are seeking an international partner, they chose among these three countries. In essence, these three large players (plus France as the fourth, and Scandinavia more recently as the fifth) compete with each other for partners in the global network. Partners that cooperate with authors in England are less likely to cooperate with the United States, for example.

4. Observations and Conclusions

We set out to explore whether the increase in international cooperation, documented in a number of articles, is affecting the organisation of science at the global level. The data suggest that, indeed, between 1990 and 2000, international co-authorship relations not only grew spectacularly in volume, but that this layer of knowledge production has become more pronounced as an interconnected structure. Expansion is shown by the increased number of countries linking into regional networks, providing evidence of a diffusion of scientific capacity to peripheral countries, and interconnection is shown by the growth in linkages between country pairs. The analysis further suggests that the network is becoming more decentralized, with regional “hubs” emerging, and a strong core group of collaborating countries growing from 6 to 8 countries.

The increased volume of internationally co-authored publications seems to have reinforced emerging structures at the global level. The global level can be considered as providing increasingly a system of reference other than the national systems. However, this system is highly structured: The factor analysis reveals that some of the leading countries compete for co-authorship relations with less developed countries. Secondary networks like the one carried by the Soviet Union and its allies have faded away during the 1990s. Although a greater number of countries are connected, some independent networks remain disconnected from the core structure in 2000.

In summary, this data show that the center-periphery model of international scientific collaboration can be replaced with a model that accounts for various centers that both collaborate among them and compete with one another for human resources from smaller national systems. The reputationally-controlled reward structure of science (Whitley, 1984) functions at the supra-national level and can thereby reinforce these developments. This supra-national level is internally differentiated, but at the same time hierarchically structured. A core group is both competitive and highly related. At the lowest levels of the hierarchy, segments of the world are not yet connected to the global structure.

References

Advisory Council on Science and Technology, 2000, Reaching Out: Canada, International Science and Technology, and the Knowledge-based Economy, the Expert Panel on Canada's Role in International Science and Technology (Industry Canada, Ottawa).

Ahlgren, Per, Bo Jarneving, and Ronald Rousseau (in preparation), Requirement for a cocitation similarity measure, with special reference to Pearson's correlation coefficient, Journal of the American Society for Information Science & Technology, 54, 550-560

Batagelj, V., & A. Mrvar, 2000, Drawing genealogies, Connections 21, 47-57.

Beaver, Donald deB., 2000, Reflections on scientific collaboration (and its study): Past, present and prospective. Keynote lecture delivered at the Second Berlin Workshop on Scientometrics and Informetrics/Collaboration in Science and in Technology held in Berlin.

Ben-David, Joseph, 1971, The scientist's role in society: A comparative study (Englewood Cliffs, New Jersey).

Crawford, Elisabeth T., Terry Shinn (editors), 1993, Denationalizing science: The contexts of international scientific practice (Kluwer Academic Publishers, Dordrecht).

Doré, Jean-Christopher, Tiiu Ojasoo, and Yoshiko Okubo, 1996, Correspondence factorial analysis of the publication patterns of 48 countries over the period 1981-1992, Journal of the American Society for Information Science 47 8, 588-602.

Franck, Georg, 2002, The scientific economy of attention: A novel approach to the collective rationality of science, Scientometrics 55 1, 3-26.

Fruchterman, T., & E. Reingold , 1991, Graph drawing by force-directed replacement, Software--Practice Experience 21, 1129-1166.

Galison, Peter, 1987, How experiments end (University of Chicago Press, Chicago).

Georghiou, Luke, 1998, Global cooperation in research, Research Policy 27, 611-626.

Gibbons, M., et al., 1994, The new production of knowledge: The dynamics of science and research in contemporary societies (Sage Publications, London).

Glänzel, Wolfgang, 2001, National characteristics in international scientific co-authorship relations, Scientometrics 51 1, 69-115.

Hamers L. et al., 1989, Similarity measures in scientometric research: The Jaccard Index versus Salton's Cosine Formula, Information Processing & Management 25 3, 315-318.

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

Katz, Sylvan and B.R. Martin, 1997, What is 'research collaboration?' Research Policy 26. 1, 1-18.

Kim, Ki-Wan, 2002, Global science, local research: The adoption and development of physics in South Korea, EASST conference paper.

Leydesdorff, Loet, 1995, The challenge of scientometrics: The development, measurement, and self-organisation of scientific communications (DSWO Press, Leiden University: Leiden); at http://www.upublish.com/books/leydesdorff-sci.htm

Leydesdorff, Loet, 2001, A sociological theory of communication: The self-organization of the knowledge-based society, (Universal Publishers: Parkland, FL); at http://www.upublish.com/books/leydesdorff.htm

Leydesdorff, Loet, and Rolf Zaal, 1988, Co-words and citations relations between document sets and environments, Informetrics 87/88, Leo Egghe and Ronald Rousseau (editors) (Elsevier: Amsterdam) 105-119.

Luukkonen, Terttu, Olle Persson, and Gunnar Sivertsen, 1992, Understanding patterns of international scientific collaboration, Science, Technology, & Human Values 17 1, 101-126.

Luukkonen, Terttu, R.J.W. Tijssen, O. Persson, G. Sivertsen, 1993, The measurement of international scientific collaboration, Scientometrics 28 1, 15-36.

Melin, G. and O. Persson, 1996, Studying research collaboration using co-authorships, Scientometrics 36 3, 363-377.

Michelet, Bertrand, 1988, L'Analyse des Associations, Thèse de doctorat (Université Paris VII: Paris).

Narin, Francis, 1991, Gobalisation of research, scholarly information, and patents – ten year trends, Proceedings of the North American serials interest group (NASIF) 6th Annual Conference, The Serials Librarian 21, Nos.2-3.

Newman, M.E.J, 2000, The Structure of Scientific Collaboration Networks, Santa Fe Institute Working Paper. Santa Fe Institute webpage publications.

Okubo, Y., J. F. Miquel, L. Frigoletto, and J.C. Doré, 1992, Structure of international collaboration in science: Typology of countries through multivariate techniques using a link indicator, Scientometrics 25 2, 321-351.

Price, D. de S., Little Science, 1963, Big Science…and Beyond, (Columbia University Press, New York).

Schott, Thomas, 1998, Ties between center and periphery in the scientific world system: accumulation of rewards, dominance and self-reliance in the center, Journal of World Systems Research 4, http://csf.colorado.edu/jwsr/.

Schott, Thomas, 1991, The world scientific community: Globality and globalization, Minerva 29, 440-462.

Schubert, A., W. Glänzel, and T. Braun, 1989, Scientometric datafiles. A comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields 1981-1985, Scientometrics 161-6, 3-478.

Shils, E., 1988, Center & periphery: an idea and its career, 1935-1987, 250-282, in L. Greenfield and M. Martin (editors) Center: Ideas and institutions (University of Chicago Press, Chicago).

Stein, Josephine, 1999, External relations in the European Union, the United States and Japan and international research and technological development cooperation, (Policy Research in Engineering Science and Technology, Manchester, United Kingdom).

Stichweh, Rudolf, 1996, Science in the system of world society, Social Science Information 35, 327-340.

Smith, David and J. Sylvan Katz, 2000, Collaborative approaches to research, HEFCE Fund review of research policy and funding, Final Report, (University of Sussex, United Kingdom).

Wagner, Caroline, Science and foreign policy: the elusive partnership, Science and Public Policy, December 2002.

Wagner, Caroline, Irene Brahmakulam, Brian Jackson, Anny Wong, Tatsuro Yoda, 2001, Science and Technology Collaboration: Building Capacity in Developing Countries? (RAND, Santa Monica).

Wagner, Caroline S., Allison Yezril, and Scott Hassell, 2000, International Cooperation in Research and Development: An Update to an Inventory of U.S. Government Spending, (RAND, Santa Monica).

Watts, D.J. and S. H. Strogatz, 1998, Collective dynamics of "small world" networks. Nature 393, 440-442.

Zitt, Michel, E. Bassecoulard, Y. Okubo, 2000, Shadows of the past in international cooperation: Collaboration profiles of the top 5 producers of science, Scientometrics 47, 627-657.

A revised version has been published as:

Caroline Wagner and Loet Leydesdorff (2005). Mapping Global Science using International Co-authorships: A Comparison of 1990 and 2000. International Journal of Technology and Globalization 1(2), 185-208.