Worlds of scientific cooperation. Choosing appropriate data for specific networks.

Myriam Baron, Denis Eckert, Marion Maisonobe

To cite this version:

Myriam Baron, Denis Eckert, Marion Maisonobe. Worlds of scientific cooperation. Choosing appropriate data for specific networks.. The 17th European Colloquium on Quantitative and Theoretical Geography, Sep 2011, Athènes, France. halshs-00825094

HAL Id: halshs-00825094
https://shs.hal.science/halshs-00825094

Submitted on 22 May 2013

WORLDS OF SCIENTIFIC COOPERATION.
CHOOSING APPROPRIATE DATA FOR SPECIFIC NETWORKS.

Myriam BARON[1], Denis ECKERT[2], Marion MAISONOBE[3]

[1]Associate Professor, UFR GHSS,UMR Geographie-cités, University Paris Diderot, Case 7001, Paris, 75205, France. Email: baron@parisgeo.cnrs.fr
[2]Researcher, UMR LISST-CIEU, University of Toulouse-Le-Mirail, Maison de la Recherche, 5 allées Antonio-Machado, 31058 Toulouse Cedex 9, France. Email: eckert@univ-tlse2.fr
[3] PhD Student, UMR LISST-CIEU, University of Toulouse-Le-Mirail, Maison de la Recherche, 5 allées Antonio-Machado, 31058 Toulouse Cedex 9, France. Email: marion.maisonobe@univ-tlse2.fr

ABSTRACT

This paper discusses data commonly used to study scientific cooperation, especially over the last three decades. Among the extensive bibliography using these data, our selection of articles highlights geographical levels most used to study "Worlds" of Scientific Cooperation in order to outline the advantages and limits of such a choice. Generally, the articles examined here do not take into account changes undergone by spatial units over time, such as modifications in weight and in the centrality index. These articles also tend to focus on key scientific fields such as nanotechnology. Very few studies examine small scientific communities such as the "DNA Transcription and Repair Group", even from a sociological viewpoint. After focusing on the publishing locations of the Group's first papers, we will define the field's trajectories over time by trying to answer questions such as "How does this small community's network evolve?" "Are the original locations still central to the evolving network?" This case study of the "Worlds" of Scientific Cooperation leads to questions such as "Are scientific activities substantially different from economic activities, such as the flow of finance or goods?" "What can a multi-level approach to Scientific Cooperation offer?"

KEYWORDS

Centralities, Changes, Database, Graph Theory, Scientific Communities

1. INTRODUCTION: FROM KNOWLEDGE ECONOMY TO WORLDS OF SCIENTIFIC COOPERATIONS

Since the early 1980s, more and more research has focused on phenomena linked to Globalization, especially corporate networks and the interaction between emblematic places such as global cities. Soon after, what is now commonly termed "Knowledge Economy" started garnering growing interest. This can be seen in the many studies on clusters developed over the past decade. Such a context leads to questioning, on the one hand, the nature of phenomena selected to reflect this knowledge economy, and on the other hand, the analytical levels that are given priority and the position occupied by a place or a system of places. The first element can be summarized thusly: the choice of phenomena depends on the period of study. For example, in the early 1980s to the late 1990s, research, particularly European projects, focused on higher education and training, placing the concept of Human Capital center stage. It was not until the late 1990s that a change was recorded in issues related to Knowledge Economy. It was at this point that certain themes were highlighted, such as the "Brain Drain" from developing countries to richer and more developed countries, or from Europe to North America. This period also highlighted the mobility of researchers and students and the policies implemented in the construction of integrated regional areas of Knowledge, such as those conducted within the Erasmus' European framework. Unfortunately, the necessary information to study such phenomena is not always available. In the early 2000s, it was possible to define, measure, and study student exchanges between European cities in the context of Erasmus programs. However, their evolution throughout that first decade is impossible to track due to insufficient data: starting in the mid-2000s, information on departure and arrival locations for Erasmus exchange students were no longer recorded.

More recently, the spread and the proliferation of university rankings worldwide represent a break in approaches to phenomena reflecting Knowledge Economy, by relegating teaching/training activities to the background in order to highlight research, emphasizing certainly its social importance but more so its economical impact through new patents, publication of major research findings, and the many signs of recognition and awards that follow. Publication data has therefore been increasingly used to identify key locations of what is sometimes called "World Science".

This article presents the different aspects of what we will call the Worlds of Scientific Cooperation and the privileged geographic patterns they reveal as deduced from the analysis of co-signatures of articles. Initially, we will present the structure of the bibliographic information used. Then, from a selection of recent articles, we will highlight approaches most often deployed to study the co-signatures of articles, and we will also indicate the places most often studied, as well as the preferred levels of analysis. Finally, to end, we will use the example of a relatively small scientific community, the DNA Transcription and Repair Group, to highlight the limitations of studies that are only interested in co-publications at the global (macro) level or in important scientific communities (nanotechnologies). Our case-study highlights not only the desirability of working on small scientific groups but also of giving the priority to a multi-level approach to scientific cooperation.

2. FROM DATA TO WORLDS OF SIENTIFIC « HAPPY FEW »

The data used here were extracted from the Web of Science (WOS), a database of scientific publications owned by Thomson-Reuters. This database contains decades of scientific literature, gradually expanding its coverage over time (number of indexed journals, the recent appearance of "proceedings"). More than 2 million records were produced in 2008. This database unevenly covers the different fields of scientific publication. The coverage of the social sciences (SSCI) and the humanities (A & HCI) are widely criticized for example. There are also problems in certain areas of engineering sciences. However, the section devoted to the hard sciences (SCI-Expanded) has better coverage in that the majority of journals related to the encompassed fields make up the perimeter of the database. In 2008, there were approximately 7000 journals for the SCI-Expanded section of the WOS database. In the case we are studying here, a rather well defined scientific community within molecular biology (DNA repair), there are no particular problems. According to the experts we consulted, the SCI-Expanded adequately covers the scope of publications by researchers in this field. This database however, should not be taken as a reflection of the total activity undertaken by researchers; it only covers a well defined and formalized segment: publications in scientific journals.

2.1. What are the elements of a WOS (Web of Science) record that interest us?

Science Studies often use WOS data to measure the progress of scientific production, co-signing relationships, how scientists quote each other, etc. This paper will use data from the following fields:
TITLE / AUTHOR / SOURCE (journal name) / ABSTRACT /KEYWORDS (subfields: Author, Keyword Plus) /ADDRESS

- The fields pertaining to article *contents* (which identify articles produced by the scientific community that we have chosen to study) are: TITLE, ABSTRACT, and both KEYWORDS fields. Among these fields, ABSTRACT and KEYWORDS have only been present since 1990-91. In this paper, we have therefore only used data posterior to 1990 (1990-2007); studying in particular co-authorship data from 1993 to 2007.
- The field "ADDRESS" (affiliation) identifies the institutional affiliation of the author(s) and the city and country of said institution. There are as many entries as there are institutions involved in the publication.
- The field AUTHORS

We used scientific contribution as a measurement unit, be it for an individual or a laboratory. An article co-written by three authors from the same laboratory counts as "1" if one considers the laboratory's production. If signed by three authors from three different laboratories (even if they are in the same city), this corresponds to 3 contributions. On the other hand, in terms of relationships between individuals (and not between places or institutions), each author counts as one contribution and establishes a link with all of his/her coauthors.

From a geographic standpoint, it is possible to encode the scientific production, in order of decreasing accuracy: by country (the easiest classification and one that has long been established by the specialists of Science Studies); by location ("City"); by institution. Thus, it is possible to analyze the number of publications by location: X number of publications in such or such place (or country). However, it is also possible to use co-authorships as a measure of scientific collaborations. One can for example build networks of individuals who co-sign articles, or build geographic networks (measuring co-authorships between institutions and/or cities...). It is this latter type of analysis that was chosen for this paper.

2.2. A partial scientific world, hierarchy or concentration?

Though the information available on the Web of Science in general and the Science Citation Index in particular, allows us to analyze snapshots of what certain dub, using a debatable expression, "*world science*" (Wagner, 2008), a key fact deserves to be emphasized: this "*world science*" is not handled exhaustively. Often, the selection is limited to only the 50 to 75 countries that contribute the most to world scientific production or the links between these countries that represent more than 10 publications. Studies therefore often focus on the 40 to 75 major cities that produce the bulk of world science (Matthiessen *et al.*, 2010; Matthiessen *et al.*, 2002). Organizing cities in terms of the number of their publications and/or their co-publications stems from a hierarchical approach. However, we know that the summit of such a hierarchy is highly stable. Finally, this hierarchical approach has a tendency to focus the debate on the issue of geographic concentration (or dispersal) of scientific production. Yet today, after 20 years of dispersal of the scientific and academic maps of a number of countries, we are witnessing the accrued presence of emerging countries such as China, Korea, Brazil, etc. The pertinence of the concept of “critical mass” as an organizing concept of national, and even supra-national scientific landscapes is greatly debated; the analysis of world science necessitates different methods than those that have been used until now. Indeed, very few studies have taken into account the fact that publications, and even more so co-publications, have increased significantly over the past 20 years. It has therefore become necessary to systematically confront the absolute and relative measures characterizing this hierarchy in order to avoid univocal and caricatural discourses on either concentration or dispersal.

The extremely hierarchical approaches to networks of co-publications must be completed by the construction of non-valued graphics. When constructing such a graphic between countries, it is possible to differentiate the centre of the network from the peripheries whose composition remains relatively stable over time. There are exceptions. Of note is China that has followed a spectacular trajectory resulting in a cooperation profile that is today almost identical to that of the United States, of Great Britain, or of France. This approach that appears to strengthen a hierarchical vision of the phenomenon also raises the question of the role of spatial proximity in the practice of co-signing articles. Analyses that take into account this last dimension, combining it with a ranking of publication locations and the densification of co-publication networks, are quite frequent in Europe (Hoekman *et al.*, 2009; Frenken, 2002). They aim to measure the evolution of cooperation within the European Union, often with the goal of evaluating “the effectiveness” of incentive policies in terms of building EU research networks. However, these studies do not take into account the different territorial levels of these policies, be they European, national, or even regional.

2.3. Studying a specific scientific community through its collaborative network: the contribution of the sociology of science

Models describing the emergence of research areas are first to be found in the sociology of science. To understand social organization processes among scientists, communication patterns were analyzed during the 1960s using sociometric analyses. In 1962, focusing on the genesis of molecular biology, Nicholas Mullins investigated the so called ‘Phage group’. He distinguishes four stages in the structuration of this group leading progressively to the emergence of ‘a scientific specialty’ (Mullins, 1962). At first, there is a paradigmatic phase from which cooperation is created between individuals sharing a common scientific interest. The resulting communication network changes into a cluster as scientific teams and their interrelations strengthen. In the end, a specialty emerges if the field is institutionalized (regular meetings, a specialized journal, students). Even though Mullins does not examine the spatial structuration of the Phage group, his model is equally pertinent to such an analysis. In our study, his model will guide the analysis of the evolution of collaboration patterns linking scientific teams working on the same research issue. Doing so, we intend to demonstrate that the diffusion and collaboration patterns of scientific knowledge differ from the flow of capital and goods. Leading cities for specific research issues do not always coincide with the 'global cities' that dominate economic geography. Moreover, the collaboration networks of the former possess specific features linked to the nature of their scientific activity. According to Diana Crane, '*If social organization exists in a research area it is of a highly elusive and relatively unstructured variety*' (Crane, 1969). As a result, co-authorship relationships linking scientific teams within a specific field are expected to be extremely disparate.

3. CASE STUDY: THE DNA REPAIR GROUP

In this section, we will focus on the evolution of a co-authorship network within an active molecular biology community. Not only will the locations of the group's papers be analyzed but also the intensity of links between these locations. Science is produced at the level of a problem area through an "invisible" social organization with unclear institutional boundaries. Diana Crane, in her analysis of the nature of such a group of scientists, explains, '*the problem area can best be understood as a temporary unit which deals with special problems and then dissolves after one or several decades when the problems have either been solved or been determined to be unsolvable*'. The scientific community studied here examines DNA repair pathways of UV damaged lesions of DNA[1]. In 1993, with the help of a team of biochemists located in Strasbourg, the DNA Repair team of Rotterdam discovered a link between the NER Pathway and the DNA transcription process. It was the beginning of a new stage of research for NER scientists. In all the publications of our sample, the words "DNA Repair" and "DNA Transcription" appear together with the name of at least one of the three diseases. The resulting sample of 1040 publications demonstrates the critical role the Dutch played in this field after the 1993 breakthrough. Considering the limited number of papers per city related to this topic we can assume the basic unit of analysis to be the team.

3.1. Myth and realities of the "Dutch Army"

In the years following the major 1993 breakthrough, NER biologists aimed at isolating the genes involved in both the NER Pathway and the DNA Transcription. From 1993 to 2000, Rotterdam was the central city of the co-authorship network both in terms of nodality (0.46) and betweenness (0.16). At the global and local level, Rotterdam controlled the structuration of this collaboration network (Figure 1). According to the American biologist JE Cleaver, a pioneer of DNA Repair research, Dutch teams were competitive in identifying multiple complementation groups, cloning the genes involved, and developing mouse models (Cleaver, 2001). These specific contributions in the 1990s implied the organization of what Thomas Kuhn calls "normal science", which is the stage of research involving "puzzle-solving" (Kuhn, 1962). In terms of social organization, Mullin's communication network results in this stage of research.

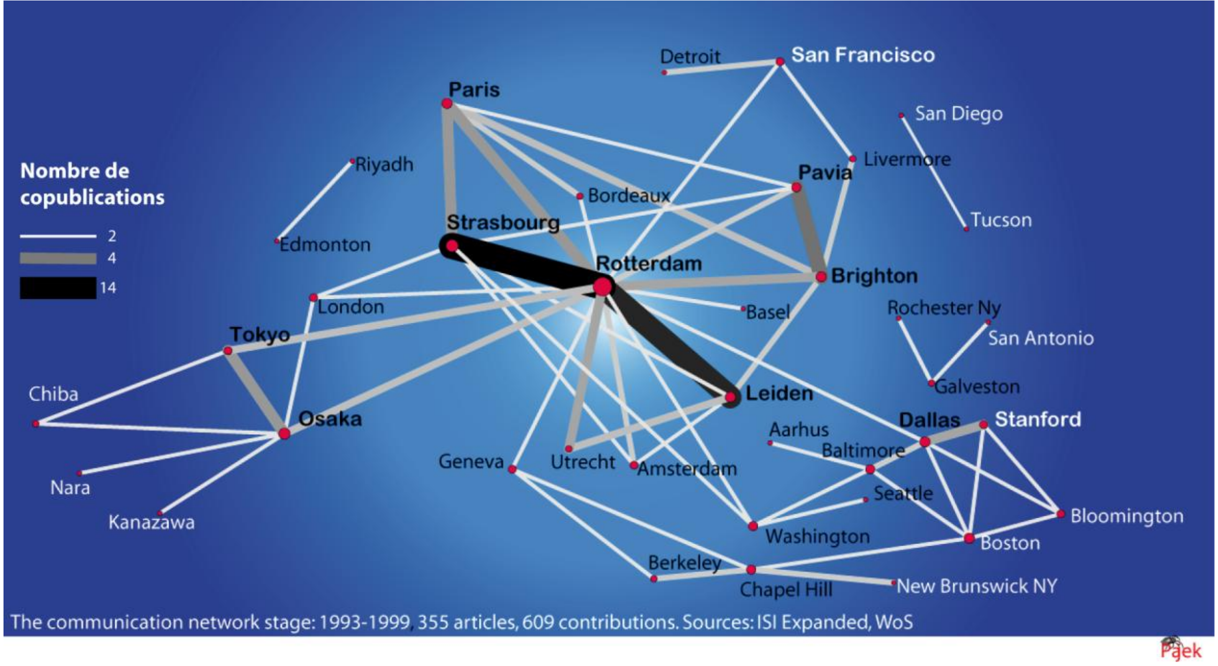
From 1993 to 1999, the so called 'Dutch Army' was characterized by a close collaboration between Rotterdam and Leiden (Figure 1, 12 articles were co-signed on this specific issue during this period). At the same time, the Rotterdam DNA Repair team had a special relationship with the Strasbourg DNA Transcription (14 articles). Moreover, Rotterdam had strong links with Brighton, Pavia, and Paris (Villejuif), which were the other European locations centralizing defective NER cells. The Rotterdam team was therefore very attractive for the other Dutch teams: Leiden, Utrecht, and Amsterdam. This explains their heightened collaboration with Rotterdam instead of each other; a phenomenon resulting from the well-known '*preferential attachment*', captured best in the phrase "*the rich get richer*".

To understand the local pattern of the 1990s, we need to look back to an earlier stage. Since the 1960s, the Dutch pioneer Dirk Bootsma had been leading a strong DNA Repair team at the Erasmus University of Rotterdam. At the time, very few teams were investigating the NER pathway, they were at what Mullins

---

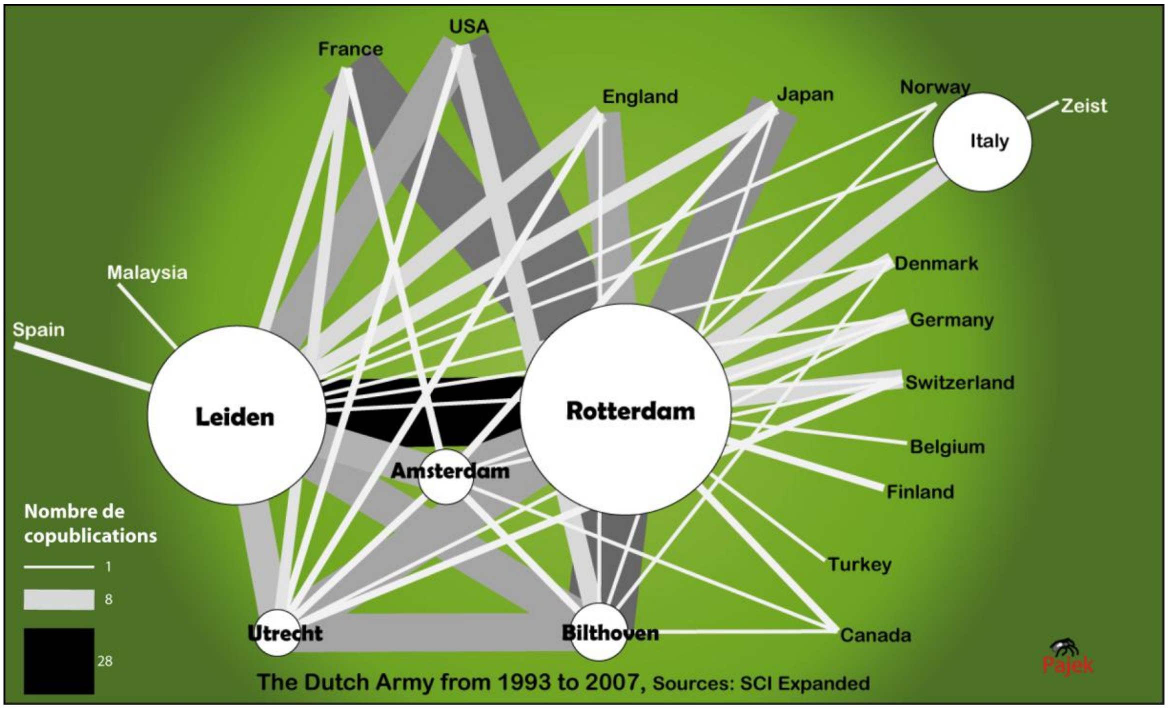
[1] This specific pathway is called the Nucleotide Excision Repair pathway or NER pathway. Since the l960s, in order to understand this repair mechanism, scientists have focused on three orphan diseases with defective NER: Xeroderma Pigmentosum (XP), Cockayne Syndrome (CS), and Thrichothiodystrophy (TTD).

refers to as the paradigmatic research stage. Bootsma’s team started collaborating with geneticists from the University of Leiden during the 1970s. The seniority of both the teams and their links explains their central role in the problem area (Bootsma, 2001). After the 1993 breakthrough, other Dutch teams got involved in the problem area. The expansion of the Dutch network began in Utrecht: the National Institute for Public Health and the Environment (Bilthoven) in collaboration with the University of Utrecht benefited from the help of Rotterdam to make mouse models. The network grew to include Amsterdam at the end of this first period (1993-1999). Thus, between the first period (P1: 1993-1999) and the second (P2: 2000-2007), the link between Rotterdam and Leiden increased by only 33% (from 12 to 16 collaborations). In contrast, Rotterdam increased links with Utrecht by 228%, tripling its previous collaboration while doubling its links with Amsterdam.



The communication network stage: 1993-1999, 355 articles, 609 contributions. Sources: ISI Expanded, WoS

Figure 1: The co-authorship network created following the 1993’s breakthrough



The Dutch Army from 1993 to 2007, Sources: SCI Expanded

Figure 2: The Dutch cluster from 1993 to 2007

However, even though Rotterdam remained central during the second phase (2000-2007), a national network was established between all the Dutch teams. By 2007, Amsterdam collaborated equally with Leiden and Rotterdam. Utrecht however still had twice as many collaborations with Rotterdam than with Leiden; this is mainly due to Bilthoven's National Research Institute (see the graph below). At the global level, Leiden's betweenness centrality index doubled (from 0.04 to 0.08) between P1 (1993-2000) and P2 (2000-2007); whereas Rotterdam's index increased be a mere 0.02 points from 0.16 to 0.18. It appears that the major changes in collaboration patterns benefitted the leaders of the field less than the "secondary" teams. The national expansion of this research network increased Netherlands' overall visibility in the DNA Repair field at the global level.

3.2. From an expanding network to the global village

From 2000 to 2007, the global co-signature network opened up to 120 new cities. These new cities co-wrote 192 publications, nearly 50% of the 403 additional publications that appeared between both periods. Whereas the top ten publishing cities of the first period (1993-1999) contributed 9% of publications, the top ten cities of the second period contributed 22% to the growth between 2000 and 2007. Even though the majority of these "historic" top ten cities maintained a high level of publication, certain second string cities caught up during the second period. This feature of the expansion process is well described by Powell (2005) in his work on Biotechnology: "*not all early entrants turn out to be winners and some latecomers attain prominence. As the saying goes, the early bird may catch the worm but it is the second mouse that eats the cheese.*" Furthermore, in terms of collaborations between the first and second periods, the betweenness centrality index, of Rotterdam and Washington remained constant whereas Paris and Osaka doubled their rates to match Rotterdam and Washington (0.15). As for the closeness centrality index, similar results can be observed: the most central cities (Paris, New York, Leiden, and Osaka) of the second period started off in peripheral positions during the first period. Conversely, certain cities previously central in both betweeness and closeness centrality indexes during the P1, lost ground during the second period P2. Apparently, what E.C Friedberg (a DNA Repair research pioneer) calls the "Texas Mafia" (Dallas, Houston, San Antonio, and Galveston) was less involved in the NER problem area from 2000 to 2007 (Friedberg, 1997). This does not mean they had stopped investigating DNA Repair after 2000; rather, it suggests they were working on other issues related to DNA Repair, as the field continued to specialize at a rapid pace. This case deserves further investigation.

Cities	**First period P1 (1993-1999)**		**Second period P2 (1993-1999)**	
	Closeness centrality index	**Betweenness centrality index**	**Closeness centrality index P2**	**Betweenness centrality index P2**
Houston	0.42	0.15	0.27	0.06
Dallas	0.42	0.10	0.39	0.02
Boston	0.41	0.07	0.39	0.04
London	0.37	0.11	0.39	0.04
Chicago	0.33	0.13	/	/

Table 1: The co-authorship network cities that lost the most ground between 1993-1999 and 2000-2007.
Sources: SCI Expanded

Since 2000, DNA Repair research has become more institutionalized. Indeed, *DNA Repair*, a specialized journal, was created in 2001 and several conferences were organized by the DNA Research community. The American "DNA Repair interest group", based in Baltimore and Bethesda (peripheral to Washington D.C.), attracted 1200 subscribers worldwide in 2005. The Baltimore and Bethesda teams are located in National Research centers founded by the National Institute of Health (NIH). Previously, both centers were located in Bethesda; but the National Institute for Aging relocated to Baltimore. In order to maintain joint training sessions and to facilitate communications between NIH centers, video-conferences have been organized since 1995. According to team leaders (Kraemer and Bohr, 2005), this virtual network has expanded to such an extent that the "DNA Repair interest group" now resembles "a global village". The group's focus problem area has played a major role in institutionalizing the DNA Repair field. Indeed, the work of "Transcription and Repair" scientists was sufficiently visible in the 1990s to attract new "members" in the following decade. Moreover, most of the highly visible scientists in the field of

"Transcription and Repair" played a major role in the organization of DNA Repair as an independent molecular biology specialty. For instance, the Baltimore based Bohr team doubled the number of their publications on "Transcription and Repair" issues between the first and the second periods of our study.
While the network continued to expand after 2000, the cognitive interest that had previously shaped the network continued to evolve. The link between Rotterdam and Strasbourg, the first period's strongest one, reduced by half during the second period. As explained earlier, their link resulted from a close collaboration following Rotterdam's original breakthrough, with the help of the Strasbourg, in 1993. Thus, this link appears to be "opportunistic", in that the different skills of each team were sufficiently complementary to be shared in furthering investigation of a common interest. However, when their investigation was complete, their link was no longer useful. Whereas Rotterdam links tended to remain fairly stable, ties between the Dutch and Japanese clusters strengthened between both periods. This suggests that '*preferential attachment'* tended to decrease while other attachment dynamics such as '*homophily'* or '*triadic closure*' were shaping the network. Indeed while the total number of Japanese and Dutch intercity links were respectively multiplied by 1.2 and 0.6 between the two periods, the total number of national links were multiplied by 2 and 1.2. Although the strategic ties of the first period did not significantly evolve, local clusters were developing. In our problem area, the growth of national ties was higher than the total growth of intercity ties for five of the six highest publishing countries (Netherlands, Japan, France, England, Italy); the exception being the USA. It would seem that the former emulating period led by the "Dutch army" led to a global network whose leaders have contributed to the structuring of national collaboration networks. Leydesdorff and Rafols have found similar evolution patterns in the network formation of two emerging technologies. They noticed that as *preferential attachment* decreases, the networks become *small-worlds* (Leydesdorff and Rafols, 2011).

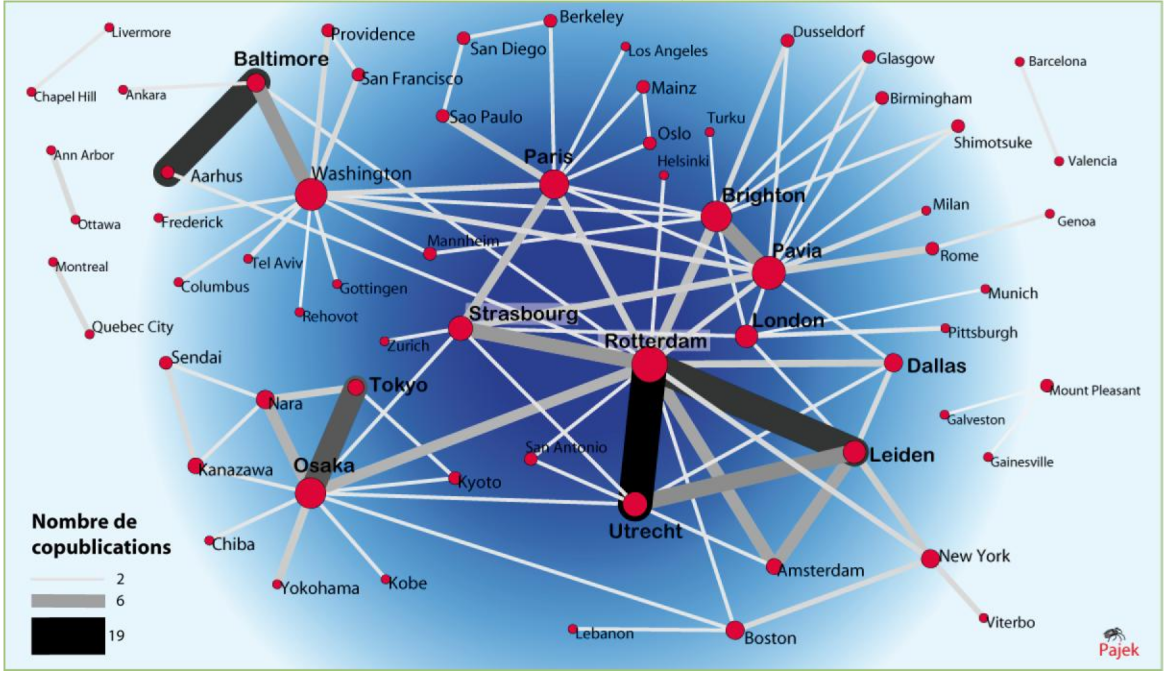


Figure 3: The 5 main centers of the "Global village": Rotterdam, Osaka, Washington, Paris, and Pavia in 2007.

4. CONCLUSIONS

This case study of the "DNA Repair Interest Group" demonstrates the necessity of increasing studies of scientific communities with limited perimeters to validate or negate the hypothesis of international structurations preceding more or less complex intra-national structurations. The multiplication of viewpoints would enrich analyses on the typologies of scientific networks, analysis that was renewed following the publication of Caroline Wagner's 2008 study. Reactions to her book have centered on the way in which the author considers the actors of scientific production as entities (researchers, laboratories) acting independently on a global stage, by effacing the role of states or integrated regional organizations (the European Union) in the management of science. This leads to an increased risk of considering scientific production as a network or networks of researchers and/or laboratories. Similarly, one can approach economy exclusively through transnational corporations and their structuration, without taking into account the organizing power of the States.

There is little to no current research that takes into account the question of building cooperation networks within a sufficiently complex, multi-level geographical approach. It is a vast undertaking referred to in the latest contribution of Mathiessen et alii (2010), which also raises the question of how to consistently define spatial objects (cities, towns, metropolitan areas, etc.), without overlooking the reliability and consistency of the different statistical corpora used.

5. REFERENCES

Bootsma, D. (2001) “The ‘Dutch DNA Repair Group’, in retrospect”, *Mutation Research*, **485**, pp. 37-41.

Cleaver, J.E., Karplus, K., Kashani-Sabet, and M., Limoli, C.L. (2001) “Nucleotide excision repair ‘a legacy of creativity’”, *Mutation Research*, **485**, pp. 23-36.

Crane, D. (1969) “Social structure in a group of scientists: a test of the ‘invisible college’ hypothesis”, *American sociological Review*, **34**(3), pp. 335-352.

Frenken, K. (2002) “A New Indicator of European Intégration and an Application to Collaboration in Scientific Research”, *Economic Systems Research*, **14**(4), pp. 345-361.

Friedberg, E.C. (1997) *Correcting the Blueprint of Life: An Historical Account of the Discovery of DNA Repair Mechanisms*, Plainview: Cold Spring Harbor Laboratory Press.

Hoekman, J., Frenken, K. and Oort (van), F. (2009) “The geography of collaborative knowledge production in Europe”, *Annals of Regional Science*, **43**, pp. 721-738.

Kraemer, K.H. and Bohr, V.A. (2005) “The DNA Repair Interest Group: a global village”, *DNA Repair*, **4**, pp. 405-406.

Kuhn, T. (1962) *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

Matthiessen, C.W., Winkel Schwarz, A. and Find, S. (2010) “World Cities of Scientific Knowledge: Systems, Networks and Potential Dynamics. An Analysis Based on Bibliometric Indicators”, *Urban Studies*, **47**(9), pp. 1879-1897.

Matthiessen, C.W., Winkel Schwarz, A. and Find, S. (2002) “The Top-level Global Research System, 1997-99: Centres, Networks and Nodality. An Analysis Based on Bibliometrics Indicators”, *Urban Studies*, **39**(5-6), pp. 903-927.

Matthiessen, C.W., Winkel Schwarz, A. (1999) “Scientific Centres in Europe: An Analysis of Research Strength and Patterns of Specialisation Based on Bibliometric Indicators”, *Urban Studies*, **36**(3), pp. 453-477.

Wagner C. (2008) *The New Invisible College. Science for Development*, Washington D.C.: Brookings Institution Press.