NBER WORKING PAPER SERIES

MEASUREMENT AND EXPLANATION OF THE INTENSITY OF CO-PUBLICATION IN SCIENTIFIC RESEARCH: AN ANALYSIS AT THE LABORATORY LEVEL

Jacques Mairesse
Laure Turner

Working Paper 11172
http://www.nber.org/papers/w11172

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2005

We wish to thank Michèle Crance, Serge Bauin and the members of the Unité des Indicateurs de la Politique Scientifique (UNIPS, CNRS), for their help in compiling the data base and their advice. We have also benefited from remarks by Marie-Laure Allain, Céline Allard-Prigent, Anne Crubelier, Dominique Foray, Claudine Hermann, and are particularly grateful for detailed comments to Bronwyn Hall and Edward Steinmueller. This paper has been written for a book (forthcoming): New Frontiers in the Economics of Innovation and New Technology: Essays in honour of Paul David, eds. C. Antonelli, D. Foray, B. Hall and E. Steinmueller, Edward Elgar Publishing. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Measurement and Explanation of the Intensity of Co-publication in Scientific Research: An Analysis at the Laboratory Level
Jacques Mairesse and Laure Turner
NBER Working Paper No. 11172
March 2005
JEL No. D29, O39

ABSTRACT

In order to study networks of collaboration between researchers, we propose a simple measure of the intensity of collaboration, which can be easily interpreted in terms of relative probability and directly aggregated at the laboratory level. We use this measure to characterize the relations of collaboration, as defined in terms of co-publication, between the physicists the French "Centre National de la Recherche Scientifique" (CNRS), in the field of condensed-matter, between 1992 and 1997, and to investigate how they vary with regards to various factors: mainly the geographical distance between laboratories, but also their specialization and size, their productivity and the quality of their publications, and their international openness. We find that the average intensity of co-publication within laboratories is about 40 times higher than the intensity between laboratories but within towns, and 100 times higher than the intensity between laboratories and between towns. Yet, geographical distance does not have a significant impact, or a very weak one, on the existence and intensity of co-publication of researchers located in different towns. We also find that the productivity laboratories, their size and proximity in specialization profiles are significant factors of collaboration.

Jacques Mairesse
CREST-INSEE
18 Boulevard Gabriel PERI
92245 MALAKOFF
CEDEX
FRANCE
and NBER
jacques.mairesse@ensae.fr

Laure Turner
ENSAE
15, Boulevard Adolphie PINARD
92245 MALAKOFF
CEDEX
FRANCE

1 Introduction

Since the effectiveness of the scientific research system has become essential in our knowledge-based economies, an important new research field has opened up. The challenge is to illuminate the role of science in economic dynamics and that of scientific institutions in the production, diffusion and transfer of knowledge. The "new economics of science" therefore analyses questions as varied as the institutional configurations of scientific systems, the job market for researchers, the incentives provided to researchers, the allocation of public funds to research, and scientific policy. It thus contributes to the understanding of the efficient organization of science (Dasgupta & David, 1994; Gibbons and al., 1994; Diamond, 1996; Stephan, 1996; Callon et Foray, 1997; Shi, 2001).

The work presented in this paper is in keeping with the focus of the economics of science on knowledge-production, and is part of a broader study of the determinants of researchers' productiveness. We believe that membership in a dynamic laboratory favors collaboration between researchers and their own individual productivity, and that it may be part of a process of cumulative advantage by which scientists enhance their productivity and recognition.[3]. Given the substantial increase in the proportion of articles co-authored by scientists who may even belong to different institutions or countries (Gibbons et al., 1994), the units of knowledge production seem increasingly to be specific networks of researchers.

---

[3] For a simulation analysis of this process in the institutional context of the US, see David (1994), and for a first attempt of an econometric analysis on the same data as one used in the present work, see Turner and Mairesse (2002).

In the economics of science, literature on the interactions that generate knowledge production primarily concerns geographic externalities that promote the local emergence of new knowledge. Authors have focused their study on the existence of such externalities within industry or between public and industrial research, basing their conclusions on patent citation data.[4] Our work, by contrast, moves further upstream to study knowledge externalities within the scientific institution by means of co-publication data. We wish to look beyond the observation of the spatial dimensions of research activity to analyze the determinants of the existence of collaboration links and of the intensity of those links. Audretsch and Stephan (1996) have contributed in this direction, yet in the framework of the relations between public research and industry. Based on data on the position of academic scientists in US biotechnology firms, they show that collaboration between firms and researchers in the same area is highly likely when the researchers' academic reputation is good, when they belong to a geographically extensive network, or when they are involved in knowledge transfer towards the firm (having participated in the creation of the firm or as a member of the Scientific Advisory Board). Regional characteristics also seem to influence the strength of the relationship between scientists and firms. In the same spirit, we wish to identify various factors that are likely to play a part in the constitution and nature of networks of collaboration between academic researchers.

A number of studies, often sociometric or bibliometric, have highlighted some of the determinants facilitating collaboration within academic research (see Katz, 1994, for a

[4] Three of these studies can be mentioned here. Jaffe (1989) shows that there is in the U. S. a close relationship at the state level between the number of patents and the importance of university research, which he interprets as evidence of geographic externalities. Jaffe, Trajtenberg and Henderson (1993) account for the localization of knowledge externalities using patent citation data. The authors show that citing and cited patents belong to the same geographic region with a very high probability. Jaffe and Trajtenberg (1998), also on the basis of patent citation data, confirm the localization of flows of knowledge on an international scale. Patents whose inventors live in the same country have a 30 to 80 percent greater chance of mutually citing one another than inventors in different countries.

summary presentation). They include, above all, researchers' reputation, popularity and visibility, demand for specific instruments needed for research, increasing specialization in science, and geographic proximity. But a rift exists within this literature, depending on the definition of the concept of a network (Shrum & Mullins, 1988). In one strand of analysis the actors in networks are identified through the relations they have between themselves. They are mainly distinguished by their position in a structured network (e.g. whether they occupy a central position or not), not by the individual characteristics and intrinsic qualities predating the place occupied in the network, such as age, gender or skills.[5] By contrast, the second line of research is based on recognition of the differing qualities of the actors with respect to status, capacities and strategies, and it is these individual characteristics that determine the position of agents in networks and the nature of interactions between them.[6]

Yet it would be desirable to include in the same analysis structural and individual elements as constituents of networks and particularly of networks of collaboration in scientific and technical research. Knowledge and innovation production and diffusion are based on the interaction of multiple agents and institutions with diverse interests: scientists in public and private laboratories, firms, financiers, public authorities, etc. (Callon, 1999).

---

[5] The theory of graphs clearly illustrates this approach since individuals are represented as inter-related "points" or lines and columns of an adjacent matrix whose coefficients express the extent of the relations. For example, by adopting a definition of networks as a set of relations exceeding a certain density threshold, called a "clique", Blau (1973) make the following observations for a group of 411 physicists. Members of large networks are often young, work in new and innovative specialities, have a teaching post and are relatively well-known; by contrast, members of small networks are older, work in established specialities, in prestigious university departments, and are involved in administration. This seems to reflect the existence of a cycle in research careers, leading the most productive scientists to be also part of the administrative elite.

[6] Some of the analysis by Cole and Cole (1973) on stratification in science is exemplary of this approach. They classify physicists in terms of different criteria such as age, prestige within university departments, productivity and scientific awards. They then measure the impact of these characteristics on the researchers' rank in the scientific system (in terms of reputation and visibility). This study is extended to the evaluation of discrimination against scientists on the basis of race, gender and religion.

Studying the structure of connections between actors by taking into account their specific characteristics should afford insight into the various mechanisms at play.

In this contribution we present the first results of an investigation carried out along these lines. We propose an intensity measure of collaboration between researchers in different locations, which has an intuitive interpretation and can be simply aggregated to the laboratory level or a higher level of aggregation. Our unit of analysis in the paper is the laboratory and the group of laboratories at the geographic level of towns. Our purpose is to explain measured differences of intensity by various factors: geographic distance between laboratories, their thematic specialization, their size, their productivity in terms of number of publications per researcher and quality in terms of citations impact factor, and their international openness.[7] In particular to what extent does the geographic distance between researchers and their laboratories strongly impede, or not, their scientific collaboration?

We identify the collaborative relations between the French physicists in the field of condensed matters of the Centre National de la Recherche Scientifique (CNRS), through their publications and co-publications during the period 1992-97.[8] Our approach is basically

---

[7] In future work, a possibility will be to extend this research at the level of individual researchers. In addition to what we already can observe at the laboratory level, this should allow the analysis of the role of productive and well-known "star" scientists in the formation and development of collaborative networks. See for example the work by Crane (1969 and 1972), Zucker, Darby, and Brewer (1998), and Zucker, Darby, and Armstrong (1998), highlighting the importance of "star scientists" in shaping research networks.

[8] The Centre National de la Recherche Scientifique (CNRS) is the main French public organization for basic research. With 25,000 employees (11,000 researchers and 14,000 engineers, technicians and administrative staff) and over 1,200 research and service units (laboratories) throughout the country, the CNRS covers all fields of knowledge. Directly administered by the Ministry responsible for research which is also usually responsible for higher education, the CNRS has very close links with the academic research pursued within the universities, with researchers from the CNRS and from universities often working in the same laboratories.

descriptive. We first characterize the extent of collaboration between these researchers, within their laboratories and across them, or within and across these laboratories aggregated at the town level, in terms of our proposed intensity measure of collaboration. We then simply assess by means of correlations the individual impact of the geographical distance and the other factors of interest a on collaboration. We then try to disentangle these different factors in estimating their relative weight in a regression analysis.

The paper is organized as follows. In section 2, we give precisions on the scope of our study, the construction of our sample and main variables, and some general characteristics of collaboration. In section 3, we define our measure of intensity of collaboration and give a detailed example of its computation. In section 4, we present our results and comment on what they tell us of the respective importance of the various factors of collaboration we have been able to consider. We briefly conclude in section 5.

2 Scope of the study and general characteristics of the collaboration

2.1 Scope of the study: collaboration between CNRS researchers in condensed matter physics

In this paper we study networks of collaboration between 493 physicists belonging to the condensed matter section at the CNRS, over the six year period 1992-1997.[9] These

[9] We have been able to follow up female researchers who married during this study period and change names.

physicists were born between 1936 and 1960 and were still working at the CNRS in 1997.[10] The physic of condensed matter was chosen for two reasons: First, the characteristics of this field are particularly well suited to our study: it a field of basic research, which is clearly defined and where the journals with a sound reputation are easily identifiable, and there is relatively very little mobility among researchers outside of the field in CNRS or out of CNRS towards academia or industry. Second, condensed matter is a fast-growing field, honored by the Nobel Prize for Physics awarded to Pierre-Gilles de Gennes in 1991, and currently accounting for close to half of all French research in physics. Condensed matter studies, at various scales (atom, molecules, colloids, particles or cells), all states of matter between liquids and solids, in which molecules are relatively close to each other. It is based on a heritage of traditions, both experimental (crystallography, diffusion of neutrons and electrons, magnetic resonance imagery, microscopy, etc.) and theoretical (solid state physics). It has recently developed a closer relation with industry, contributing to the development of materials used in electronics, composites, plastics, food or cosmetic gels, and so forth.

The group of 493 physicists studied here represents a majority of all CNRS researchers in this discipline. The CNRS and higher education institutions are the only public research institutions in this domain in France. In 1996, there were a total of 654 condensed matter physicists in CNRS, as against 1475 in universities and “Grandes Ecoles” (Barré, Crance, and Sigogneau, 1999).

[10] This criterion for the selection of researchers was mostly based on two practical considerations: they had to be "not too young" so that we had a history of their publications (the youngest researchers born in 1960 had already been publishing for a few years in 1992, when they were 32 years old); secondly, 1997 was the year for which we could know precisely the laboratories in which the researchers were working, when we first started compiling our data base.

The fact that our study is limited to researchers belonging to the same institution, the CNRS, comes in fact as an advantage. It generates a strong organizational proximity between them, characterized by the sharing of common knowledge and implicit or explicit rules of organization that favor interaction and coordination (Rallet and Torre, 2000; Foray, 2000). Because they all belong to the same scientific community in the same institution , they work in a context conducive to informal cooperation, that is, cooperation that does not involve the prior definition of rules of coordination. The existence of this organizational proximity thus makes it possible to isolate more clearly the effects of geographic distance proper on collaboration.

The indicator of collaboration that we use in this study is co-publication. It seems to be a reliable indicator of collaboration without being an exhaustive measurement, in so far as collaboration can have results other than publication. Our data base has been compiled on the basis of all the publications drawn from the Science Citation Index (SCI), for 518 CNRS condensed matter physicists over the period 1992-1997, of whom 493 published at least one co-authored article during this six years. [11] Of the remaining 25 physicists, 21 in fact publish no articles in this period, and the other 4 published only a total of 5 non-co-authored articles. Collaboration appears to be the main mode of publication for the 493 researchers. Only 132 of them also wrote articles without co-authors over the period (for a total of 252 articles), and from the total corpus of 7,784 articles they wrote over the period, 7,532 (97%!) are in fact co-authored.

[11] The *Science Citation Index* (SCI) is produced by the Institute for Scientific Information (ISI). It encompasses all the (hard) scientific disciplines and is constructed on the base of a compilation of over 3,200 of the most cited international periodicals. The quality of the data is remarkable and, in particular, the coverage of scientific publications by CNRS units is very satisfactory (UNIPS, 1999). Ninety-five percent of the scientific articles written by the CNRS researchers are in English and these are fully covered by the SCI.

In so far as we wished to qualify the intensity of collaboration, we thought appropriate in our analysis to give more weight to an article in proportion of the numbers of coauthors it involves. In other words we simply chose to study the network of collaboration "link by link," that is, by pairs of authors. In practice, this means that an article features in the computation of our measure on collaboration intensity analysis in the data-base as many times as it concerns different pairs of authors.[12].

We also chose to base our study here at the aggregate level of the laboratory, and even at the level of the group of laboratories in the same town or locality. We thus consider networks of collaboration between laboratories or groups of laboratories ("towns") rather than directly between individual researchers. When two researchers belonging to different entities collaborate, we consider that their entities collaborate, and on this basis we can measure the intensity of collaboration between pairs of laboratories and towns. When two researchers belonging to the same entity collaborate, we also simply consider it as a case of collaboration "within" this entity, and likewise we compute the intensity of collaboration within laboratories and within towns. We can also similarly compute intensity of collaboration between laboratories-within towns. Carrying our study at the aggregated laboratory or town level makes the illustration of our measure of collaboration intensity somewhat more easy, the network of collaboration being much denser at this level than at the individual researcher level. But it also has the advantage of allowing to characterize directly the influence on collaboration of working in the same laboratory or locality, and thus telling us about the importance of immediate proximity and easiness of the face-to-face relations.

[12] Another solution would be to count each article only one time, by simply weighting them by the inverse of the number of pairs of authors concerned. This point is discussed in section 3.3. It seems that the main results of our analysis would have been qualitatively unchanged.

2.2 Two configurations of collaboration

The coauthors of the articles of 493 CNRS researchers, which we will simply call "CNRS researchers" from now on, can also be (these) CNRS researchers themselves, or other researchers from universities or other institutions, either French or foreign, which we will call "external researchers". In our analysis, we are led to distinguish between these two categories of researchers and to separate two configurations of collaborations, depending on whether a publication involves *at least two CNRS researchers* and possibly other researchers (CNRS or external), or whether it concerns *at most one CNRS researcher* and others external researchers. A basic reason for this distinction is a matter of practicality. CNRS researchers were the only ones for whom we knew the exact address of their laboratories.[13] To identify the location of the external researchers we would have had to be able to determine it reliably from the SCI, which was not possible.[14] But without the location of the "others," we were unable to measure the geography of collaborative relations between them and a CNRS researcher. Therefore we will consider primarily the case of collaboration involving at least two CNRS researcher in the following discussion.

A total of 1,823 articles corresponds to the first form of collaboration (at least two CNRS researchers and others), and 5,709 articles to the second form of collaboration (at the

[13] We obtained this information directly from the Unité des Indicateurs de la Politique Scientifique (UNIPS) of CNRS.

[14] In the SCI the number of authors recorded for an article is rarely equal to the number of addresses listed, and no key for correspondence between authors and addresses exists. It is, for example, possible that several authors out of all those collaborating on one article have the same address, in which case the address will appear only once on the list. But when the collaboration also involves other laboratories, we cannot know to which author to attribute the address. Another frequent example is that of multiple signatures, that is, one author who signs her/his affiliation to several laboratories, so that the number of addresses becomes greater than the number of authors, resulting again in a problem of attribution.

most one CNRS researcher and others). The 493 physicists who collaborate do so most often in these two modes. However, 38 physicists collaborate only with CNRS researchers on our list and never with others, and 69 researchers collaborate only with others and never with another CNRS researcher on our list. The 1,823 articles corresponding to the first mode of collaboration – *"Group 1"* – involve 424 of our researchers (= 493 – 69), while the 5,709 articles in the second mode – *"Group 2"* – involve 455 researchers (= 493 – 38). The selection of the sample is summarized in Figure 1.

[Figure 1 about here]

Table 1 shows the distribution of the number of articles of Groups 1 and 2 according to the number of their authors, depending on whether those belong to the CNRS or not. The immediate observation is that for both groups collaboration involves a large number of "other" researchers i.e. non CNRS (only 82 articles are written by CNRS researchers only). Thus, for an average of 5.9 authors per article for Group 1 and 4.9 for Group 2, the average number of "other" researchers per article is 3.7 and 3.9 respectively. We also note that an article rarely involves more than two CNRS researchers, as the third line of Table 1 shows. Only 20% of the articles in Group 1 are co-authored by more than two CNRS researchers, which corresponds to an average of 2.2 CNRS authors per article in Group 1.

[Table 1 about here]

2.3 The selected sample and some other characteristics of collaboration

Our paper relies on the *Group 1* sample, containing the 1,823 articles whose authors consisted of at least two CNRS researchers on our list. Four main reasons determine this choice, of which two have already been mentioned. The first is of a practical order. Since we do not have the location of "other" researchers, we cannot geographically locate the collaboration defined by articles in Group 2 (at the most one CNRS researcher among the authors). The second reason is analytical. By studying articles in Group 1, we study collaboration of pairs of CNRS researchers. We thus control for the organizational proximity created by "common knowledge" of practices and know-how of the institution, which can lead to coordination even without geographic proximity or any other contextual factor. This enables us to isolate the impact of geographic proximity on collaboration.

But there is a third reason for selecting Group 1 articles. The density of relations of co-publication between researchers must be high enough to allow the study of collaborative networks. From this point of view, the number of collaborative relations per pair corresponding to articles in Group 1 is greater than the number of relations corresponding to Group 2. For Group 2, the 5,709 articles involve 455 CNRS researchers and close to 10,000 others, a total of 17,500 pairs with at most one CNRS researcher. By contrast, in Group 1, the 1,823 articles have been written by 424 CNRS researchers and about 3,500 others, or by 880 pairs with at least two CNRS researchers. Thus, the average number of articles per pair of authors is only 0.33 for Group 2 as opposed to 2.1 for Group 1.

The last element explaining the restriction of the study to Group 1 articles is the fact that Groups 1 and 2 have comparable characteristics, which suggests that the results of a study performed on all the data would not be very different from those obtained by studying the first group only. In order to show this, we limit ourselves to the articles situated at the intersection of the two groups, that is, the 6,753 articles published by the 386 CNRS researchers who collaborate in both modes.[15] The first shared characteristic is the frequency of the number of articles in relation to the number of "other" co-authors, as shown in Figure 2. Thus, the probability that an article is co-authored by a particular number of "other" researchers is the same in both groups. The second shared characteristic is the degree of concentration of the number of articles, as shown in Figure 3. For both groups of articles the concentration curves are very similar, which shows that the inequality of the productivity distribution is similar in the two modes of collaboration. In particular, in both cases 30% of the articles are written by 10% of the most productive researchers.

[Figures 2 and 3 about here]

Yet the number of articles per researcher differs somewhat in the two groups. During the period under study, a researcher publishing in Group 1 (i.e. at least two CNRS researchers) wrote an average of 9.9 articles, while a researcher publishing in the second mode of collaboration (with "others") wrote an average of 13 articles, that is, 30% more per annum[16]. Figure 4 shows the cumulative distribution of productivity of the 386 researchers according to the two forms of collaboration. The cumulative probability that a researcher publishes less than six articles in Group 1 is 50% and is greater than the cumulative probability that a researcher publishes less than six articles in Group 2 (35%).

---

[15] Subtract from the 493 researchers who collaborate the 38 who never publish with "others" and the 69 who never publish with other CNRS researchers to obtain 386. See Figure 1.

[16] Note, this mean does not correspond to the simple mean, calculated as the ratio between the number of articles (1,823) and the number of researchers (386). The idea is to count each article as many times as there are CNRS authors, in order to attribute to each author her/his stock of publications. The sum of

[**Figure 4**]

We operated a last treatment of our data. Notice that the 493 scientists who collaborate are geographically concentrated in two regions around Grenoble and Paris (more than 60% of the researchers locate in these regions). We decided to restrict our sample in order to keep laboratory and towns of sufficient size in terms of the number of scientists hosted. We imposed that towns had at least nine researchers and laboratory at least five researchers. Our final sample consists then of 470 scientists (out of 493) located in 17 towns and 34 laboratories. To simplify the analysis, only pairs of towns (resp. laboratories) with more than 6 (resp. 4) collaborations are assigned a value for the count of collaborations; for those with fewer collaborations, the count is set to zero. This did not change the number of towns nor laboratory in the sample, but the number of articles considered in group 1: the total number of group 1 selected articles is 1634. To sum up, we are studying a sample of 470 scientists located in 17 towns and 34 laboratories, and are concentrating on 1634 group 1 articles (i.e. associating at least two of those scientists as authors).

3 Measurement of intensity

Inclusion of agents in networks is determined by "intrinsic" individual qualities such as age, gender, skills and strategy, and by more structural variables such as number of relations they develop, geographic distance, etc. As a result, the form and functioning of networks differ. If the actors were not differentiated and if they collaborated with all the others with equal probability, we would expect to observe a uniform structure of relations

---

individual stocks of publications is 3,813 in group 1. We thus have a weighted mean of CNRS articles between all the individuals. We take this case of homogeneity as a reference. At the aggregate level of laboratories and groups of laboratories (towns) on which this work is based, the case of homogeneity corresponds to a configuration in which the frequency of collaboration of each entity with all the others is the same, irrespective of their geographic localization and of the characteristics of the entities.

Our novel measure of intensity of collaboration between two entities is simply based on the comparison between the real network described by the data and the network that would be observed in the case of perfect homogeneity of the entities. Section 3.1 defines this measure, which is commented in section 3.2 and 3.3. Then an practical example of its calculation is provided in section 3.4.

3.1 Definition

In this section we assume for simplicity that collaboration always involves at the most two CNRS researchers.[17] The network studied has a finite number of entities. It contains N researchers who can form C collaboration pairs (where by definition $C = N(N-1)/2$, the total number of possible combinations of pairs). The network is complete when the observed number of pairs is equal to C. Let n be the total number of articles produced in collaboration between the N researchers, then p the frequency of the number of co-publications per pair in the complete network is the ratio between the total number of articles, n, and the number of possible pairs, C.

We let the same type of notation at the level of the network's members. Consider two entities X and Y in the network. N_X researchers work in entity X and N_Y researchers in entity

---

per author given by 3,813/386 = 9.9 on average.

[17] This point is discussed in the next section.

Y. The number of possible pairs within X is $C_X = N_X(N_X-1)/2$, and within Y is $C_Y = N_Y(N_Y-1)/2$, and the number of possible pairs that can be formed between researchers from X and Y is $C_{XY} = N_X N_y /2$. Likewise, the total number of articles written jointly within X and Y is respectively n_X and n_Y, and the total number of articles written in common by researchers in X and Y is n_{XY}. The frequency of collaboration p_{XY} is the ratio between the total number of articles written in common by researchers in X with researchers in Y, and the number of possible pairs of researchers from the two entities. Similarly, the frequency of collaboration within the entity X (or Y), noted as p_X (or p_Y), is the ratio between the total number of articles written together by researchers from X (or Y), and the number of possible pairs of researchers in entity X (or Y). We have:

$$p_X = \frac{n_X}{C_X}, \quad p_Y = \frac{n_Y}{C_Y}, \quad \text{and} \quad p_{XY} = \frac{n_{XY}}{C_{XY}}$$

The intensity of collaboration relates the frequencies obtained at the entities' level to the frequency p obtained for the complete network. We define *intra-* and *inter*-entity intensities as follows:

Intra-entity intensity: $i_J = \frac{p_J}{p}$, with $J = X$ *or* Y.

Inter-entity intensity: $i_{XY} = \frac{p_{XY}}{p}$

In case of perfect homogeneity of the network we have $p_X = p_Y = p_{XY}$ for all X and Y. Consequently in that case : $p_X = p_Y = p_{XY} = p$ or in terms of the intensity measure: $i_X = i_Y = i_{XY} = 1$ (this comes simply from adding up the numerators and denominators for each entity). So if the links are homogeneous in distribution, the frequency for the complete network p is the frequency that we expect to find for X and for Y. This is because we aggregate equal probabilities of collaboration between individuals. And if the entities of the studied network are homogeneous, the intra- and inter- intensities of collaboration are all equal to unity.

Otherwise – and this is the case with a real network described by data like ours – various factors influence the collaboration intensities so that measured intensities generally do not equal 1. Therefore in commenting our results in section 4 we take 1 as a benchmark.

Another way of looking at the measure of intensity is to rewrite it as the contribution of a unit to the total number of articles published in collaboration in the network (called *n*), normalized by the relative size of the unit in the complete network in terms of possible pairs of collaborators. In case of homogeneity, when intensity equals 1, the contribution in terms of co-published papers of a unit is in exact proportion of its relative size in the network. We have: $i_J = \frac{p_J}{p} = \frac{N_X / n}{C_X / C}$ and $i_{XY} = \frac{p_{XY}}{p} = \frac{N_{XY} / n}{C_{XY} / C}$.

The structure of intensity of a network of *E* entities can be represented by means of a symmetrical matrix *E* by *E* with coefficients that are either positive or zero, and of which the diagonal terms are intra-entity intensities, and the lines (or columns) inter-entity intensities.[18] Appendix 1 shows such a matrix for the 17 towns in our sample.

3.2 Properties of aggregation

Intensity as defined above has the advantage of being easy to aggregate. In order to see this, take *X* and *Y*, the two laboratories in town *V*. The total number of co-authored articles written in *V* is the sum of co-authored articles written in *X*, in *Y*, and by the pair of

[18] Note that this matrix is similar to the adjacency matrix used in the graph theory. The coefficients of the adjacency matrix are equal to 1 when there is a link between the entities represented on the lines and those represented in the columns; otherwise it is 0. The adjacency matrix is a representation of indicators of inter- or inter-entity collaboration that does not take the number of authors per article nor the strength of the ties into account. By contrast, the matrix of intensities provides more qualitative information on the collaboration.

laboratories *X* and *Y* jointly. Likewise, the number of theoretical pairs of researchers in *V* is the sum of the theoretical pairs of researchers in laboratory *X*, in laboratory *Y*, and between the two. We have:

$$\frac{n_V}{C_V} = \frac{n_X + n_Y + n_{XY}}{C_X + C_Y + C_{XY}}$$

This formula can also be written as follows:

$$\frac{n_V}{C_V} = \frac{n_X}{C_X} x \frac{C_X}{C_V} + \frac{n_Y}{C_Y} x \frac{C_Y}{C_V} + \frac{n_{XY}}{C_{XY}} x \frac{C_{XY}}{C_V}$$

$$or$$

$$p_V = p_X \left(\frac{C_X}{C_V} \right) + p_Y \left(\frac{C_Y}{C_V} \right) + p_{XY} \left(\frac{C_{XY}}{C_V} \right)$$

We thus have a formula of aggregation that can generalize to a network of laboratories. By dividing the formula by *p*, we obtain the intensity of collaboration of a network as the weighted sum of intra- and inter-entity intensities. The weightings represent the weight of each laboratory in the network in terms of possible collaboration pairs. Aggregating over the entire network, we have

$$p = \sum_I p_I \frac{C_I}{C} + \sum_{I,J \neq I} p_{IJ} \frac{C_{IJ}}{C}$$

$$\text{or } 1 = \sum_I \frac{p_I}{p} \frac{C_I}{C} + \sum_{I,J \neq I} \frac{p_{IJ}}{p} \frac{C_{IJ}}{C}$$

$$\Rightarrow 1 = \sum_I i_I w_I + \sum_{I,J \neq I} i_{IJ} w_J$$

where

$$w_I = \frac{C_I}{C} \qquad w_{IJ} = \frac{C_{IJ}}{C} \qquad \text{with} \qquad \sum_{I,J} w_{IJ} = 1$$

3.3 Remark on the weighting

Until now, we have supposed that the articles were written by two researchers at the most. In reality, they were also written by threesomes, foursomes, etc. But, as indicated, we consider only pairs of collaboration, and the data base repeats an article as many times as there are pairs of different authors. In other words, the total number of co-authored articles is a number weighted by the number of pairs that contribute to its publication. By counting the number of articles written by threesomes, foursomes, etc., we would have obtained the total number (not weighted) of co-authored articles in the network. For example, for an article published by three CNRS researchers, one belonging to an entity *X* and the two others to an entity *Y*, we would have counted only one collaborative link between *X* and *Y*, which amounts to counting the article only once, if we had reasoned in terms of threesomes. Instead, when we enumerate pairs of collaboration we count three links – two between *X* and *Y* and one within *Y*. This amounts to counting the article three times, as many times as the number of pairs that actually contributed.[19] This procedure is necessary for the aggregation formula to remain valid in the case where more than two researchers co-author an article. Moreover in our case, since only 20% of the articles in Group 1 are co-authored by more than two CNRS researchers, which correspond to an average of 2.2 CNRS authors per article in Group 1, the weighting should not make a large difference.

[19] We could have considered counting one third of the link three times, that is, breaking up the article into the number of pairs, so as to count the article only once. But in so far as we wished to highlight the density of collaborative relations, it seemed appropriate to consider that the more authors it has, the greater the weight of an article.

3.4 Practical calculation: an example

Let us take the concrete example of Marseille to describe the calculation of the intensity of collaboration, with reference to Table 2 which also presents the results of this calculation for the other towns studied. Marseille is a town with 18 physicists on our list (column 1). There are 34 collaborations developed within the town (column 3) and 18 with other towns (column 4), of which ten are with Grenoble and eight with Strasbourg. Marseille also has relations with Poitiers, Gif-sur-Yvette, Orsay, Toulouse and Villeurbanne, but these are not taken into account because they all involve less than six collaborations (see 2.3). The number of possible pairs of researchers working in Marseille is 18*17/2, or 153. The frequency of the number of collaborations per pair of researchers in Marseille is therefore 34/153 or 0.22.

Given that the number of researchers in Grenoble and Strasbourg is 105 and 14, respectively, the number of possible pairs of researchers linking Marseille and Grenoble is 1,890 (105*9), and linking Marseille and Strasbourg is 252 (14*18). The frequency of the number of collaborations per pair between Marseille and Grenoble is therefore 0.0053 (= 10/1,890) and between Marseille and Strasbourg 0.0317 (= 8/252).

In order to obtain the intensities of collaboration, we have to calculate the frequency of collaboration per pair for the set of the 17 towns. It is the ratio between the total number of weighted articles, 2,480, and the number of possible pairs that can be formed by the 470 researchers in that set, i.e. 110,215 pairs (470*469/2). We thus have p = 0.0225. In case of homogeneity, p is the frequency that would have been obtained for intra- and inter-Marseille's collaboration. In reality, the intra-frequency is much higher (0.22) and the frequency of collaboration between Marseille and Grenoble is lower (0.0053). The frequency of collaboration between Marseille and Strasbourg is closer to the reference value (0.0317).

The within town intensity for Marseille is then 9.88 (column 5). The intensity between Marseille and Grenoble is 0.24, and between Marseille and Strasbourg 1.41, yielding a mean intensity of collaboration between Marseille and its partners of 0.82 (column 7). In column 6 we have the mean intensity of collaboration between Marseille and all the other 16 towns (0.1).

[Table 2 about here]

4 Results

In this section we first comment the results of the calculation of the intensities at the town level (4.1). Then we present the statistical influence of various factors on the establishment of links and intensity of relations in the networks of collaboration (4.2). Finally we give our results on the determinants of collaboration obtained in a regression model (4.3).

4.1 Intensity of collaboration in networks of French towns

The results of the calculation of intensities at the town level show large differences in inter-town intensities, some towns like Paris or Grenoble being linked to almost every other towns whereas towns like Poitiers, Orléans and Talence are isolated[20]. We now comment Table 2 into more details.

[20] We recall the reader that to simplify the analysis only pairs of towns and laboratories with more than five collaborations are assigned a value for the count of collaborations; for those with fewer collaborations, the count is set to zero.

The second column gives the number of partners in the collaboration, for each of the 17 towns. This number is 4, on average, which seems rather low for towns grouping at least 9 scientists[21]. Nine out of the 17 towns have less than 4 partners among which the three towns mentioned, Poitiers, Orléans and Talence, have no significant relations with the others. Yet among those 9 towns, 3 are of relatively great size (>14 scientists). In the same way, it is remarkable that of the 136 possible pairs of towns, only 34 effectively collaborate.

The towns that have the more links are Grenoble, Paris and Orsay. They are also the largest ones. They operate as a node: 12 towns are collaborating with Grenoble (70% of the towns), 9 with Orsay (53%) and 7 with Paris (41%).

The intra-town intensities of collaboration are presented in column 5 of table 2. They are high, always greater than one, and on average equal to 18.9. The number of intra-town collaborations is then almost 19 times higher than what we would have found in case of homogeneity, which shows the importance of proximity in collaboration.

Note that towns with few partners like Meudon, Poitiers, Strasbourg, Villeneuve-d'Ascq et Villeurbanne, have very high levels of intra-intensity, as opposed to nodes like Grenoble, Orsay and Paris, for which intra-intensities are among the lowest (5.4, 3.6 and 3.0 respectively). This result could in part be explained by both the size of the scientific community in Grenoble, Orsay and Paris and the fact that they host several laboratories: large size and multiple labs entail many potential links among which relatively many do not form. As a matter of fact, at the laboratory level intra-intensities for Grenoble, Orsay and Paris are comparable with the others intra-laboratory intensities (intra-lab average intensities are 19.4, 33.4, and 34.5 in Grenoble, Orsay and Paris respectively).

[21] But we do not have other studies to compare this result with.

The intensity of collaboration between the 17 towns in the sample is presented in columns 6 and 7 of Table 2. Column 7 indicates the mean intensity of collaboration of each town with its effective partners, and column 6 presents the mean intensity of collaboration of each town with the other 16 towns. The mean intensity of collaboration of each town with its partners is low (1.01 on average), and the intensity of collaboration with all the other towns is 0.25 on average and almost always lower than 1. On the whole, towns collaborate less with one another than what one would expect in the case of homogeneity.

Moreover, collaboration between towns is far less intense than collaboration within towns, indicating the strong influence of proximity on the intensity of collaboration. This observation is confirmed when looking the intensities at the laboratory level (table 6): the average intensity of co-publication within laboratory is about 40 times higher than the intensity between laboratories but within town, and 100 times higher than the intensity between laboratories and between towns.

4.2 Determinants of the intensity of collaboration

This section presents a statistical approach to the determination of factors influencing the intensity of relations in networks of collaboration. We first implemented the approach on the level of towns and then on that of laboratories. Six determining factors are studied: geographic distance, specialization, size of the scientific community, productivity, quality of publications, and openness towards international collaboration. Descriptive statistics for these variables are given in Table 3.

[Table 3 about here]

Table 4 gives the results obtained using towns as the unit of observation. It shows the correlation between each factor studied with three indicators of collaboration: *the intensity of intra-town* collaboration (column 1), the *existence of collaboration links* between towns (column 2), and the *inter-towns-intensity* of collaboration (column 3). Defining a single measure of a characteristic for a pair of towns is problematic and we therefore used three definitions: the minimum or maximum value of the characteristic across the pair, or the average of its values for the partners in the pair.

[Table 4]

Figures 5 to 10 complement Table 4 by showing the intensity of collaboration between towns and partners in relation to the different characteristics of the pairs studied. In each figure we present the most significant relation of the three between the intensity of collaboration and the characteristic of the pairs. Thus the characteristic shown is the average of the characteristics, the minimum, or the maximum, depending on the figure.

The results obtained using the laboratory as the unit of observation are presented in Tables 5a and 5b, where we distinguish them according to whether the laboratories belong to the same town or not. Table 5a presents the correlation of these factors with the variable indicating the *existence of collaboration* between laboratories, depending on whether or not the laboratories are in the same town. The first column in Table 5b indicates the correlation of the different factors with the *intensity of collaboration within* laboratory, while the second and third describe the correlation of the factors with the *intensity of collaboration between laboratories* that actually collaborate, according to whether the laboratories are in the same town or not. In general, the results for towns and laboratories are similar, with the notable exception of the correlation with geographic distance and with specialization.

[**Tables 5a and 5b**]

We now comment these tables.

4.2.1 Distance

At the town level, the labs' readiness to collaborate with distant partners varies. Table 3 indicates that the average distance of a town from its partners can vary widely (the distance may be three times more in some cases than in others). For example, Montpellier collaborates with towns over a distance of 520 km on average, while Gif-sur-Yvette has relations with towns close by, situated at an average distance of 170 km. Four of the five towns situated less than 300 km from their partners are in the Parisian region (consisting of six towns in which our physicists are present).

Geographic distance plays no part in the *establishment* of collaboration between towns, and has no impact on the *intensity* of *inter-town* collaboration (the coefficients are not statistically significant). Thus, it is not because towns are far apart that they are less likely to collaborate.

Yet when we scale down the study to the laboratory level, we observe a negative relation between distance and the *existence* of collaborative relations between labs[22]. This is mainly due to the high values of the intensity of collaboration inter-labs within town (table 2). Again, immediate proximity favors collaboration. We see here that beyond the perimeter of the town the effect of distance is slightly unfavorable to the establishment of collaborative relations (table 5a). This confirm table 6 that shows that the mean intensity of collaboration

[22] The correlation between distance and the indicator of existing links between labs no matter the towns (not in table 5a) is equal to –0.18***.

within the same laboratory is greater than the mean intensity of collaboration between laboratories situated in the same town, and that it itself is greater than the intensity of collaboration between laboratories in different towns. Concerning the *intensity* of relations, the distance does not have a statistically significant impact on inter-laboratories intensity.

There seems to be two types of distance: immediate proximity that favors face-to-face interaction between researchers, and distance that eventually shrinks with the development of communication technologies. In prior studies on knowledge flows between public laboratories and industry, proximity is linked to the potential for face-to-face interactions which is assumed to be more effective for the exchange of tacit knowledge (where the actors must mutually build a common understanding rather than being able to refer to a common 'text') (Zucker, Darby and Armstrong, 1998). When relations are no longer face-to-face, distance is no longer a relevant factor in choosing collaborators. New communication technologies have certainly helped to reduce the role of geographic distance either by facilitating the codification of tacit know-how and allowing its diffusion (consulting data bases, reading working papers, sending articles and data, etc.) or by assisting researchers to more fluidly interact as they build shared understanding (e.g. e-mail and forms of 'chat' messaging). [23]

4.2.2 Specialization

We have tried to take into account the specialization of laboratories. The map of France presented in Appendix 2 suggests the influence of specialization on the geography of networks of co-publication. While distance plays a very small part in the intensity of

[23] It would be interesting to know whether the recent period has favoured collaboration across distance more than preceding periods. If so, a possible interpretation may be that new communication technologies have made it possible to establish new collaborative relations between distant researchers.

relations, it seems that this is also because collaborations are governed by this specialization of laboratories. In particular, storage rings,[24] which are very large facilities used by physicists of condensed matter, are present at only two laboratories, Orsay and Grenoble, which consequently appear to be central poles in the collaboration.

There are no “typical” specialization data traditionally used in research. We have defined a profile of specialization of entities, based on the main theoretical and/or experimental sub-domains of the discipline to which their journals belong. Classification of journals in these sub-domains is relatively well identified; it is carried out by the SCI. During the period 1992-97, the main sub-domains in which the entities under study published were physics-chemistry, general physics, solid-state physics, applied physics, materials science, and crystallography. We compute the vector of proportions of publications in each of these six main sub-domains, and in the other sub-domains grouped under the label “other,” and we call this vector the “specialization profile of an entity”.

[24] Storage rings are used to curve or oscillate the trajectory of light charged particles (electrons or positrons) that then emit "synchrotron radiation". This constitutes an extraordinary source of radiation of varying wavelengths, especially X-rays, and has become of great practical importance. Several rings have been built throughout the world for synchrotron radiation, the most recent of which have a circumference of about 500 meters. The USA has about ten rings and France has the "Super-ACO" and "DCI" rings situated at Orsay (at the LURE). The LURE has a total of 50 different experimental apparatus available for most of synchrotron radiation applications. About forty can work simultaneously. About 30 outside laboratories collaborate on a permanent basis with the LUREe, as do 20 industrial partners, in the field of physics but also chemistry, biology and environmental science, micro-production, lithography and astrophysics. The LURE rings will soon be replaced by the "SOLEIL" ring (in 2005), that will constitute a sort of "super" synchrotron radiation, that is, several thousand times brighter, and will thus afford possibilities for new applications in many scientific areas. The facility will be located at Saint-Aubin, near Orsay. The European ring of the ESRF (European Synchrotron Radiation Facility), owned by 16 countries, is situated at Grenoble and employs about 500 persons on a permanent basis.

The idea here is to see to what extent two entities with the same specialization profile are most likely to collaborate. For this purpose we measured the "proximity" of specialization profiles of entities two by two, and computed the correlation of this measurement with the probability of collaboration and with its intensity. To that end we used the Chi-squared statistic that allows the simultaneous comparison of several distributions of frequencies.

Secondly, we wanted to distinguish those main sub-domains that favored collaboration most. We therefore broke down the specialization profile of each entity into seven sub-profiles. Each of these sub-profiles corresponded to the degree of specialization of the entity in one of the sub-domains of the discipline or in the category "other". We again used the Chi-squared measurement to calculate the proximity between each sub-profile of two entities.

Table 3 gives the average distance, in terms of specialization, between each town and its partners, measured by means of the Chi-squared statistic. To simplify the interpretation, this measurement was normalized by the theoretical value of the Chi-squared statistic.[25] It is therefore at most equal to unity when the towns have the same specialization profile, and increases with the distance in terms of specialization between the towns. This measurement can be interpreted in the following way: the higher the value, the more similar the profiles of specialization of the two entities. Thus, if common specialization favors collaboration between the two entities, we can expect to see positive correlations between "specialization" and, respectively the indicator of the existence of collaborative relations and the intensity of the collaboration. Bagneux and Villeneuve D'Ascq are therefore towns that collaborate most with entities with a different specialization profile (the normalized measurement is roughly

---

[25] We used the value from the tabulated chi-squared distribution at the five per cent level of significance and with degrees of freedom consistent with our number of series (17) and sub-domains (7) respectively.

3), unlike Grenoble, Marseille and Talence (where the measurement is roughly 11). On average, the (normalized) proximity between partner towns with respect to their specialization is 6.1.

Tables 4, 5a and 5b present the computed correlations. The first line of the section “Specialization” corresponds to the Chi-squared calculated for the specialization profile of all the entities. The following lines include the Chi-squared calculated for each of the seven sub-profiles of the entities. At the town level, common specialization has no effect on the intensity of collaborative relations. Other results are less clear-cut.

At the laboratory level, proximity in specialization favors the establishment of collaborative relations between laboratories in the same town. A laboratory with the same overall profile and *a fortiori* the same sub-profile as a laboratory situated within its town, is more likely to collaborate with it (Table 5a). In the case where the laboratories belong to different towns, results are less clear-cut and less significant statistically. By contrast, in all the sub-domains[26] common specialization impacts negatively on the intensity of collaboration in those cases where the laboratories that collaborate are not in the same town (Table 5b). To sum up, at the laboratory level, proximity in specialization favors the establishment of collaboration links but has a negative impact on the intensity of collaboration between labs in different town. Yet we think that we must go deeper into the study of this variable before assessing these results.

4.2.3 Size

Three indicators of size are used. The first is the number of researchers present in the towns studied. The size of the scientific community is unequal in the different towns and this is reflected in our sample (Table 3). Researchers are concentrated in and around Grenoble and Paris (20% and 17% of all researchers, respectively). The second indicator of size, particularly well-suited to the definition of our measurement of intensity, is the number of possible pairs of researchers in the towns under consideration. The last indicator of size is the stock of publications between 1992 and 1997 in the towns studied. In this respect, Grenoble is the main town because it accounts for 26.5% of the total stock of publications. Paris is second, with 14% and Orsay third, with 13% (Table 3).

Tables 4, 5a and 5b indicate the correlations of the different indicators of size with the indicator of collaboration and the intensities of collaboration. They indicate that, the greater the size of the entity (town or laboratory), the more the entity develops collaborative relations with the other entities (Tables 4 and 5a, column 2). But the larger the entities the weaker the intensity of relations between them (especially when in a different town), as revealed by the negative sign of correlation between the size and intensity of collaboration between partner entities (Table 5b) and also by Figure 5. Moreover, size is negatively correlated with intensity within laboratory (Table 5b, column 1).

4.2.4 Productivity

The productivity of researchers in an entity is defined as the ratio between the stock of publications[27] between 1992 and 1997 and the number of researchers in the entity. As shown in Table 3, towns are not equally productive since the number of articles per researcher in the six towns ranges from 6.3 in Orleans to 36.4 in Bagneux. The mean is 15.7.

---

[26] Yet not significantly statistically for physics-chemistry and applied physics.

[27] All publications, including those with only one author or which belong to Group 2 of co-publications. Moreover, each article is counted only once, irrespective of the number of authors.

Collaboration intensity within entities is strongly and significantly correlated with productivity. In the case of towns, the correlation is strong using the indicator “Min productivity of I and J” but not using “Max productivity of I and J”. This result seems intuitive: a collaboration pair is formed on condition that the two towns satisfy a certain minimum level of productivity. Once this collaboration is established, the intensity increases along with the productiveness of the towns, as represented in Figure 7. In the case of laboratories, the probability of collaborative relations developing and the intensity of these relations is positively correlated to the productivity of the laboratories, without any minimum threshold constraint.

4.2.5 Quality of publications

A measure of the quality of the stock of publications by researchers in an entity is the mean of the scores given for the impact of the journals in which the articles appeared. The impact score of a journal is equivalent to the average citation rate of its articles and therefore gives information on the journal's reputation and visibility. Using this quality measurement, citations are recorded over a period of two years. Table 3 shows that the towns studied publish in journals whose articles are cited an average of three times in two years. The quality of publications is variable, depending on the town. It is lowest for Poitiers, where the average citation rate of articles in journals is 2.34, and highest for Palaiseau where the average citation rate is 4.77.

The quality of publications does not impact the within-unit intensity of collaboration. In the case of inter-town and inter-laboratory collaboration, as for productivity, the correlation of the maximum of the qualities with the variable indicating collaboration is not significant, whereas the correlation with the minimum is. This indicates that the two entities must have a stock of publications of a certain quality for collaboration to be stimulated. Yet inter-laboratory collaboration within the same town is always less probable when the quality of publications by either laboratory is high. Concerning the intensity of collaboration between entities, it seems to be independent of publication quality (as shown in Figure 8 for towns).

4.2.6 International openness

We recall that it is impossible to attribute an address to those researchers called "others." It is interesting to note what this term "others" covers, and especially to see how it refers to countries other than France. We therefore built an indicator of the presence of names of foreign countries in the variable containing addresses of the laboratories of the authors of the articles studied. With a margin of error related to the non-standardization of the variable "address" in the SCI, we thus obtained information on the proportion of articles involving foreign collaboration. The last column in Table 3 shows the proportion of articles written with foreign collaboration for each town. On average, towns collaborate with other countries in about 30% of their articles. Paris tops the list (50%) while Bagneux has the least cooperation with foreigners (12%).

We interpret the proportion of articles involving a foreign country as an indicator of international openness of the laboratories and towns under study. We therefore calculate the correlations for this variable. Openness towards foreign countries favors the existence of collaborative relations at the town level. But it has no significant impact on the intensity of intra-and inter-town collaboration (Tables 4). Yet it has a positive impact on inter-labs inter-towns intensity of collaboration.

4.3 Regression results for the laboratories

In this section of the paper we present a series of linear regressions in order to assess the robustness of the relations revealed by the statistical correlations at the laboratory level. The results are shown in Table 7, in three panels. The first panel shows the results of a linear probability model for the existence of collaboration within and between towns, the second shows the regression model of the intensity of the collaborations within and between towns, and the third shows the regression model for the intensity of collaboration within a laboratory. The factors that appear to significantly foster collaboration are the size, the productivity of the laboratories, and their localization.

The probability that two laboratories collaborate increases with their size. But the impact is relatively low: if the size of each laboratory was doubled, it would result in a number of links between labs 2.8% higher in the "inter towns" case, and 10% higher in the "intra town" case. On the contrary, the intensity of collaboration between town diminishes with the size of the partner labs, and this effect is important. In fact, if the size of each lab was doubled, the collaborative intensity between labs in different towns would be 30% lower. However, because the intensity of collaboration between laboratories located within the same town is not affected by the size, changing the size of the laboratory would not affect the intensity of collaboration between laboratories in the same town. Within lab, size has a strong negative impact on the intensity of collaboration among researchers.

In all cases, the laboratories' productivity enhances the probability that two laboratories collaborate as well as their collaborative intensity. This effect has the same order of magnitude than the size effect. If the productivity of each laboratory was doubled, it would result in a number of links between labs 2.6% higher in the "inter towns" case, and 8% higher in the "intra town" case, and the collaborative intensity between labs would be 30% higher.

The geographic distance between laboratories has a small negative impact on the probability that they collaborate, but does not influence the intensity of collaboration. For instance, if the distance between each laboratories was doubled, the number of links inter laboratories would be 0.8% lower. The proximity in specialization does not have a strong impact on collaboration and we think that we must go deeper into the study of this variable before assessing its results.

5 Conclusion

This study aimed both at proposing an measurement of the intensity of collaboration in networks and at quantifying the impact of several factors on the shape and collaborative intensity of scientific networks. The measure of the collaborative intensity that we proposed allowed us to identify the proximity of the researchers within the same entity, the size of the laboratories, and their productivity as the main determinants of the collaboration and of the collaborative intensity on a laboratory scale. Geographic proximity in the form of "face-to-face" interactions (within labs and towns) enhances the probability that researchers co-publish. But beyond immediate proximity, distance does not play a significant role in the choice of collaborators. By contrast, productivity and size have an important role in stimulating collaboration linkages between distant labs.

As was mentioned, future work could usefully extend this analysis to examine the contribution of researchers' individual characteristics (such as age, gender, promotion, reputation) to the construction of networks. It will lead us to study the "star" scientists and their role in the elaboration of collaboration links. Further development of econometric models integrating these individual factors as well as the structural determinants identified in the present paper would provide measures of their respective contributions to networks of collaboration.

Bibliography

AUDRETSCH D., FELDMAN M., 1996, "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 86(3), 630-640.

AUDRETSCH D., STEPHAN P., 1996, "Company-Scientist Locational Links: The case of biotechnology," *American Economic Review*, 86(3), 641-652.

BARRE R., CRANCE M., SIGOGNEAU A., 1999, "La Recherche Scientifique Française : Situation démographique," *Etudes et dossiers de l'OST*, 1.

BLAU J., 1973, "Patterns of Communication among Theoretical High Energy Physicists," *Sociometry*, 37, 391-406.

BEAVER D., ROSEN R., 1978, "Studies in Scientific Collaboration. Part 1. The Professional Origins of Scientific Co-authorship," *Scientometrics*, 1, 65-84.

BEAVER D., ROSEN R., 1979, "Studies in Scientific Collaboration. Part 2. Scientific Co-authorship, Research Productivity and Visibility in The French Scientific Elite," *Scientometrics*, 1, 133-149.

CALLON M., 1999, "Le Réseau comme Forme Emergente et comme Modalité de Coordination : le Cas des Interactions Stratégiques entre Firmes Industrielles et Laboratoires Académiques," in Callon M. et al., *Réseau et coordination*, Economica.

CALLON M. et FORAY D., 1997, "Introduction : Nouvelle économie de la science ou socio-économie de la recherche scientifique ?" *Revue d'Economie Industrielle*, 79, 13-37.

COLE J. and COLE S., 1973, *Social Stratification in Science*, Chicago: U. of Chicago Press.

CRANE D., 1969, "Social Structure in a Group of Scientists: A Test of the Invisible College Hypothesis," *American Sociological Review*, 34, 335-352.

CRANE D., 1972, *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*, Chicago: U. of Chicago Press.

CRAWFORD S., 1971, "Informal Communication among Scientists in Sleep Research," *Journal of the American Society for Information Science*, 22, 301-310.

DASGUPTA P., DAVID P. A., 1994, "Toward a New Economics of Science," *Research Policy*, 23(5), 487-521.

DAVID P., 1994, "A Science Simulation for Studying the US and similar Institutional Setups (SCISIMUS)," Mimeo.

DIAMOND A., 1996, "The Economics of Science," *Knowledge and Policy*, 9 (2-3).

FORAY D., 2000, "L' Economie de la Connaissance", La Découverte.
GIBBONS M. et al., 1994, *The New Production of Knowledge*, Sage.

JAFFE A., 1989, "Real Effects of Academic Research," *American Economic Review*, 79(5), 957-970.

JAFFE A., TRAJTENBERG M., 1998, "International Knowledge Flows: evidence from patent citations," NBER Working Paper 6507.

JAFFE A., TRAJTENBERG M., HENDERSON R., 1993, "Geographic Localization of Knowledge Spillovers as Evidenced from Patent Citations," *Quarterly Journal of Economics*, 108(3), 557-598.

KATZ J.S, 1994, "Geographical Proximity and Scientific Collaboration", *Scientometrics*, 31(1), p. 31-43.

LEAMER E.., STORPER M., 2000, "The Economics of Geography at the Internet Age" mimeo, University of California at Los Angeles and Management and Public Policy, November.

RALLET A., TORRE A., 2000, "Is Geographical Proximity Necessary in the Innovation Networks in the Era of Global Economy?" mimeo, Université Paris Dauphine et Institut National de la Recherche Agronomique.

SHI Y., 2001, "*The Economics of Scientific Knowledge,*" Edward Elgar.

STEPHAN P., 1996, "The Economics of Science," *Journal of Economic Literature*, 34, 1199-1235.

SHRUM W., MULLINS N., 1988, "Network Analysis in the Study of Science and Technology," in *Handbook of quantitative studies of science and technology*, ed A.F.J. Van Raan, North Holland.

TURNER L., MAIRESSE J., 2002, "Explaining Individual Productivity Differences in Public Research: How Important Are Non-Individual Determinants? An Econometric Analysis of French Physicists (1986-1997)," Working Paper- Cahiers de la MSE 2002-66, Université Paris I- Panthéon-Sorbonne.

UNIPS, 1999, "Les Publications des Laboratoires du CNRS et leur Impact," éd. CNRS, mars.

ZUCKER L, DARBY M., ARMSTRONG J., 1998, "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises," *American Economic Review*, 88, 290-306.

ZUCKER L, DARBY M., BREWER M. 1998, "Geographically Localized Knowledge: Spillovers or Markets?" *Economic Inquiry*, 36, 65-86.

Table 1. ***Number of co-authored articles published by groups 1 and 2 by number and type of authors***

Number of "non-CNRS" co-authors:	0	1	2	3	4	5	6	7	8	9	Total
Group 1 (at least 2 CNRS co-authors) Of which:	82 (38)	230	324	375	260	209	127	81	56	161	1823 (424***)
2 CNRS co-authors	64	196	268	300	218	172	106	61	47	66	1498
3 CNRS co-authors	15	31	45	60	31	33	15	15	6	7	257
4 CNRS co-authors	3	2	9	12	8	4	6	4	3	4	55
5 or more CNRS co-authors	0	1	2	3	3	0	0	1	0	3	13
Group 2 (only 1 CNRS co-author)	*	726	1087	1114	976	708	441	241	128	288	5709 (455**)
Total (group 1 and 2)	**82**	**956**	**1411**	**1489**	**1236**	**917**	**568**	**322**	**184**	**367**	**7532 (493)**

( )The numbers in parentheses below the numbers of co-authored articles are the numbers of CNRS researchers coauthoring these articles.
* The number of articles with a single author is 252. They are published by 132 scientists who also published co-authored papers.
** Including 69 researchers who never published with another CNRS scientist and who account for 697 publications of group 1 co-authored papers.
*** Including 38 researchers who never published with others and who account for 82 publications of group 2 co-authored papers.

Table 2. ***Average intra- and inter-town intensity of collaboration***

	Number of scientists	Number of partner towns (*)	Number of articles "intra"	Number of articles "inter" (*)	Intensity intra-town	Intensity inter-town (average computed on all other 16 towns)	Intensity inter-town (average computed on partner towns only)
Bagneux	9	6	51	171	63.0	1,3	3,5
Poitiers	11	0	31	0	25.1	0.0	0.0
Gif sur Yvette	16	3	11	40	4.1	0.2	0.9
Grenoble	105	12	666	449	5.4	0.7	0.9
Marseille	18	2	34	18	9.9	0.1	0.8
Meudon	9	2	27	19	33.3	0.1	0.5
Montpellier	20	7	47	83	11.0	0.3	0.8
Orléans	10	0	7	0	6.9	0.0	0.0
Orsay	66	9	174	192	3.6	0.2	0.4
Palaiseau	18	4	15	45	4.4	0.2	0.9
Paris	86	7	249	148	3.0	0.3	0.6
Saint Martin d'Hères	31	5	161	193	15.4	0.3	0.9
Strasbourg	14	2	72	20	35.2	0.1	0.9
Talence	9	0	8	0	9.9	0.0	0.0
Toulouse	29	4	88	63	9.6	0.4	1.6
Villeneuve d'Ascq	10	3	39	31	38.5	0.6	3.0
Villeurbanne	9	2	35	58	43.2	0.2	1.4
Total	**470**	**34** [d]	**1715** [a]	**765** [b]			
Moyenne		**4** [c]			**18.9** [c]	**0.3** [c]	**1.0** [c]
p	-	-	-	-		0.0225	

* Pairs of towns linked by fewer than five articles have been excluded.
a Each article is weighted by the number of pairs of authors that contribute to its publication, otherwise the number of articles would be 1222.
b Each article is weighted by the number of pairs of authors that contribute to its publication, otherwise the number of articles would be 412.
c The mean of the variable is shown.

Table 3. *Descriptive statistics for the variables*

Town	Number of laboratories per town	Number of scientists	Theoretical number of pairs “inter“	Theoretical number of pairs ”intra’”	Stock of publications between 1992 and 1997	Mean geographic distance to partners	Mean distance to partners in terms of specialization	Mean productivity	Mean quality of the publications	Proportion of articles coauthored by foreigners
Bagneux	1	9	4149	36	328	344	3.02	36.44	3.68	0.12
Poitiers	1	11	5049	55	88	0	5.17	8.00	2.34	0.26
Gif sur Yvette	1	16	7264	120	246	171	3.93	15.38	3.07	0.14
Grenoble	6	105	38325	5460	1870	421	10.19	17.81	3.39	0.50
Marseille	1	18	8136	153	235	361	11.17	13.06	2.84	0.44
Meudon	1	9	4149	36	99	208	5.97	11.00	2.63	0.20
Montpellier	3	20	9000	190	365	548	5.63	18.25	3.47	0.21
Orléans	1	10	4600	45	63	0	8.24	6.30	3.54	0.32
Orsay	3	66	26664	2145	922	334	4.97	13.97	3.69	0.25
Palaiseau	2	18	8136	153	274	297	4.32	15.22	4.77	0.33
Paris	6	86	33024	3655	985	291	6.72	11.45	3.75	0.48
Saint Martin d'Hères	2	31	13609	465	438	418	4.44	14.13	3.78	0.27
Strasbourg	1	14	6384	91	248	449	6.45	17.71	3.69	0.16
Talence	1	9	4149	36	193	0	11.07	21.44	3.94	0.43
Toulouse	2	29	12789	406	379	410	5.30	13.07	2.71	0.24
Villeneuve d'Ascq	1	10	4600	45	184	305	3.12	18.40	4.13	0.28
Villeurbanne	1	9	4149	36	139	188	4.55	15.44	3.02	0.23
Total	**34**	**470**	**194176**	**13127**	**7056**	**279.12***	**6.13***	**15.71***	**3.44***	**0.29***

*The mean of the variable is shown.

Table 4. ***Correlations at the town level***

Correlation between	Intensity within town, i_I (N=17)	(0 or 1) (N=136)	$i_{IJ} \neq 0$ (N=34)
Distance between towns	-	-0.09	-0.16
Specialization			
General Profile	-	-0.02	-0.21
Physics-Chemistry		-0.18**	-0.14
General Physics		-0.20***	-0.20
Solid-state Physics		-0.14	-0.24
Applied Physics		0.06	-0.08
Materials Science		-0.02	-0.06
Crystallography		0.16*	-0.17
Other		-0.00	-0.16
Size of the scientific community			
number of researchers			
N_I	- 0.48 *	-	-
Maximum (N_I, N_J)	-	0.52***	- 0.41**
Minimum (N_I, N_J)		0.42***	- 0.31*
Average $(N_I + N_J)/2$		0.56***	- 0.45***
number of possible pairs of researchers in the towns under consideration $C_{IJ} = N_I * N_J$	- 0.39	0.49***	- 0.32*
stock of publications between 1992 and 1997			
S_I	- 0.29	-	-
Maximum (S_I,S_J)	-	0.52***	-0.29*
Minimum (S_I,S_J)		0.54***	-0.21
Average $(S_I + S_J)/2$		0.60***	-0.33*
Productivity between 1992 and 1997			
P_I	0.62***	-	-
Maximum (P_I,P_J)	-	0.11	0.67***
Minimum (P_I,P_J)		0.26***	0.47**
Average $(P_I + P_J)/2$		0.19***	0.69***
Quality of publications			
Q_I	- 0.11	-	-
Maximum (Q_I,Q_J)	-	0.03	0.09
Minimum (Q_I,Q_J)		0.23***	0.03
Average $(Q_I + Q_J)/2$		0.16*	0.08
International Openness			
pe_I	-0.37	-	-
Maximum (pe_I, pe_J)	-	0.14*	-0.26
Minimum (pe_I, pe_J)	-	0.34***	-0.12
Average $(pe_I + pe_J)/2$	-	0.26***	-0.22

Correlations shown are significant at the 10% level (*), 5% level (**), and 1% level (***) respectively.
a: Existence of a collaboration link inter-towns. b: Intensity of collaboration between partner towns *I* and *J*

Table 5a. *Correlations at the laboratory level*

Correlation between	Existence of a collaborative link between laboratories	
	In different towns (N=522)	In the same town (N=39)
Distance between towns	-0.09**	-
Specialization		
General Profile	-0.02	0.40***
Physics-Chemistry	-0.05	0.30***
General Physics	-0.09***	0.35***
Solid-state Physics	-0.04	0.24**
Applied Physics	-0.01	0.05
Materials Science	0.05*	0.41***
Crystallography	0.06**	0.10
Other	-0.03	-0.11
Size of the scientific community		
Number of researchers		
N_I	-	-
Maximum (N_I, N_J)	0.21***	0.45***
Minimum (N_I, N_J)	0.22***	0.57***
Average $(N_I + N_J)/2$	0.24***	0.57***
number of possible pairs of researchers in the towns under consideration $C_{IJ} = N_I * N_J$	0.27***	0.61***
stock of publications between 1992 and 1997		
S_I	-	-
Maximum (S_I,S_J)	0.23***	0.58***
Minimum (S_I,S_J)	0.35***	0.63***
Average $(S_I + S_J)/2$	0.31***	0.63***
Productivity between 1992 and 1997		
P_I	-	-
Maximum (P_I,P_J)	0.19***	0.35***
Minimum (P_I,P_J)	0.30***	0.40***
Average $(P_I + P_J)/2$	0.26***	0.39***
Quality of publications		
Q_I	-	-
Maximum (Q_I,Q_J)	-0.03	-0.46***
Minimum (Q_I,Q_J)	0.06**	-0.30***
Average $(Q_I + Q_J)/2$	0.02	-0.44***
International Openness		
pe_I	-	-
Maximum (pe_I, pe_J)	0.02	0.07
Minimum (pe_I, pe_J)	0.03	-0.06
Average $(pe_I + pe_J)/2$	0.03	0.01

Correlations shown are significant at the 10% level (*), 5% level (**), and 1% level (***) respectively.

Table 5b. ***Correlations at the laboratory level***

Correlation between	Intensity within labs, i_I(N=34)	Intensity of collaboration between partner labs	
		In different towns (N=41)	In the same town (N=15)
Distance	-	-0.06	-
Specialization			
General Profile		-0.32***	-0.42**
Physics-Chemistry		-0.18	-0.10
General Physics		-0.29***	-0.30
Solid-state Physics		-0.25**	-0.05
Applied Physics		-0.05	-0.26
Materials Science		-0.21*	-0.35*
Crystallography		-0.26**	-0.16
Other		-0.11	-0.26
Size of the scientific community			
Number of researchers			
N_I	- 0.42**	-	-
Maximum (N_I, N_J)		-0.52***	-0.34*
Minimum (N_I, N_J)		-0.51***	-0.02
Average $(N_I + N_J)/2$		-0.60***	-0.27
number of possible pairs of researchers in the towns under consideration $C_{IJ} = N_I * N_J$	- 0.34**	-0.49***	-0.23
stock of publications between 1992 and 1997			
S_I	- 0.06	-	-
Maximum (S_I,S_J)	-	-0.08	0.18
Minimum (S_I,S_J)		-0.14	0.13
Average $(S_I + S_J)/2$		-0.12	0.17
Productivity between 1992 and 1997			
P_I	0.36**	-	-
Maximum (P_I,P_J)	-	0.51***	0.33*
Minimum (P_I,P_J)		0.42***	0.36**
Average $(P_I + P_J)/2$		0.54***	0.36**
Quality of publications			
Q_I	0.20	-	-
Maximum (Q_I,Q_J)	-	-0.03	-0.25
Minimum (Q_I,Q_J)		0.12	0.11
Average $(Q_I + Q_J)/2$		0.05	-0.006
International Openness			
pe_I	0.14	-	-
Maximum (pe_I, pe_J)	-	0.34***	-0.005
Minimum (pe_I, pe_J)	-	0.17	0.35*
Average $(pe_I + pe_J)/2$	-	0.29***	0.22

Correlations shown are significant at the 10% level (*), 5% level (**), and 1% level (***) respectively.

Table 6. *Intensity of collaboration and geographic distance*

Town	Number of laboratories per town	Intensity of collaboration within laboratory	Intensity of collaboration inter-laboratory and within- town*	Intensity of collaboration inter-laboratory and inter- town*
Bagneux	1	58.4	-	2.0
Futuroscope	1	23.2	-	0.0
Gif sur Yvette	1	9.2	-	0.4
Grenoble	6	19.4	2.7	0.5
Marseille	1	11.6	-	0.1
Meudon	1	30.9	-	0.1
Montpellier	3	28.1	0.0	0.3
Orléans	1	6.4	-	0.0
Orsay	3	33.4	1.4	0.2
Palaiseau	2	11.9	0.0	0.2
Paris	6	34.5	0.2	0.2
Saint Martin d'Hères	2	37.1	0.0	0.4
Strasbourg	1	38.0	-	0.1
Talence	1	15.7	-	0.0
Toulouse	2	25.7	1.5	0.2
Villeneuve d'Ascq	1	98.9	-	0.5
Villeurbanne	1	40.0	-	0.4
Total/Average	**34**	**30.7**	**0.8**	**0.3**

* Average intensity computed over all laboratories

Table 7 ***Regression results for laboratories***

Variables	Linear probability model for the existence of collaboration between laboratories			
	Within town (N=39)		Between towns (N=522)	
Constant	-0.130 (1.637)	-1.012 (0.224)	-0.290 (0.181)	-0.262 (0.588)
Distance	-	-	-0.008** (0.004)	-0.008** (0.004)
Specialization	0.023 (0.036)		-0.009* (0.005)	-0.008* (0.005)
Size	0.040*** (0.015)	0.048*** (0.007)	0.015*** (0.001)	0.014*** (0.002)
Productivity	0.031* (0.017)	0.040** (0.011)	0.013*** (0.003)	0.013*** (0.002)
Quality of publications	-0.176*** (0.021)		-0.017 (0.029)	
International openness	-0.076 (1.545)		0.167 (0.223)	
Adjusted r-squared	0.374	0.413	0.107	0.108

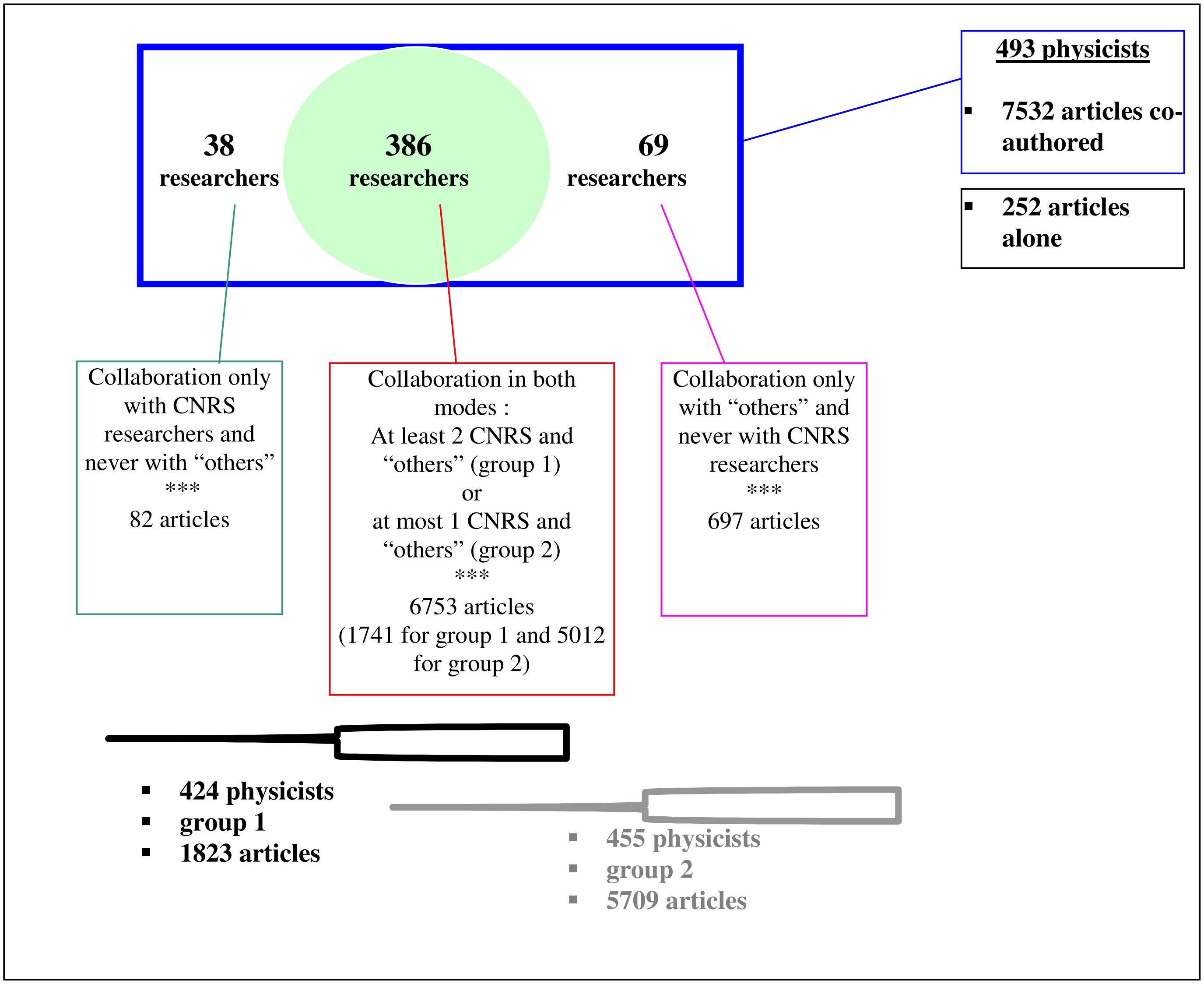
Variables	Dependent variable: intensity of collaboration between laboratories			
	Within town (N=15)		Between towns (N=41)	
Constant	-20.5 (20.85)		4.62 (11.01)	10.42 (1.346)
Distance	-		-0.02 (0.19)	
Specialization	-0.778* (0.426)	-0.499* (0.258)	-0.410* (0.226)	-0.370* (0.189)
Size	-0.307 (0.338)		-0.238** (0.091)	-0.300*** (0.066)
Productivity	0.325 (0.192)	0.344** (0.173)	0.188 (0.122)	0.312*** (0.078)
Quality of publications	-0.829 (2.417)		0.478 (1.818)	
International openness	33.36 (29.824)		-0.93 (12.13)	
Adjusted r-squared	0.162	0.280	0.369	0.385

Table 7 ***(continued)***

Variables	Dependent variable: intensity of collaboration within laboratory (N=34)	
Constant	29.105 (3.680)	29.105 (3.680)
Distance	-	-
Size	-1.507*** (0.498)	-1.426*** (0.474)
Productivity	3.314*** (1.088)	1.889*** (0.711)
Quality of publications	6.882 (5.415)	
International openness	40.028 (58.484)	
Adjusted r-squared	0.312	0.285

Significant at the 10% level (*), 5% level (**), and 1% level (***) respectively

Figure 1. *Choosing the sample*



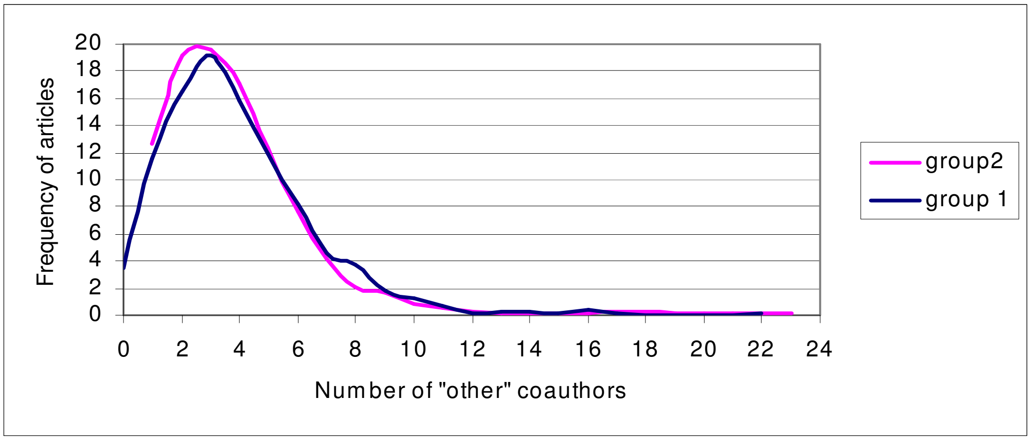


Figure 2. ***Frequency of the number of articles written with "other" co-authors***

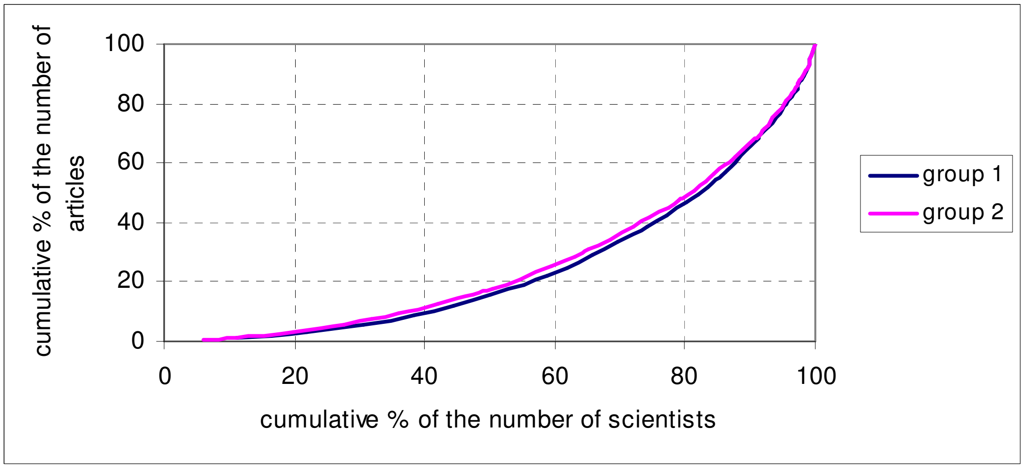


Figure 3. ***The degree of concentration of the number of articles***

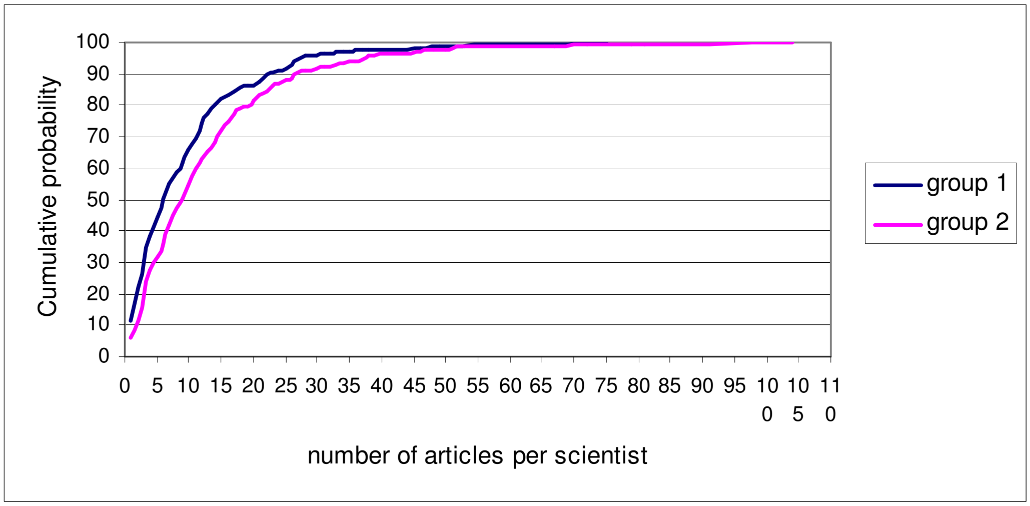


Figure 4. ***Distribution of productivity according to the two forms of collaboration***

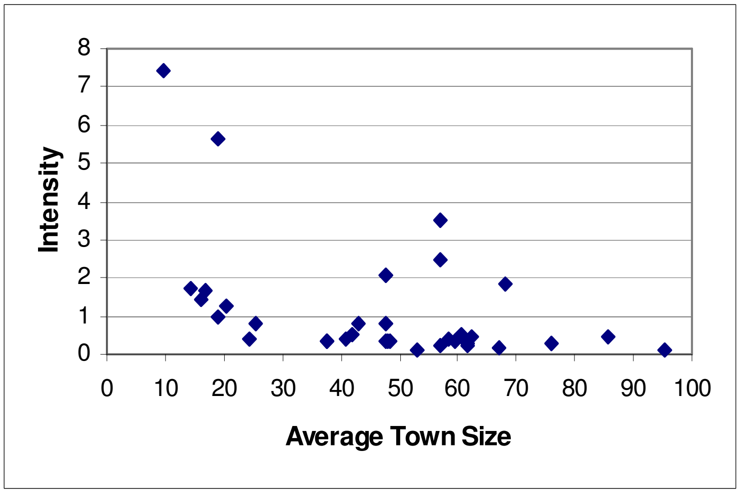


Figure 5. *Intensity of collaboration between towns versus their average size*

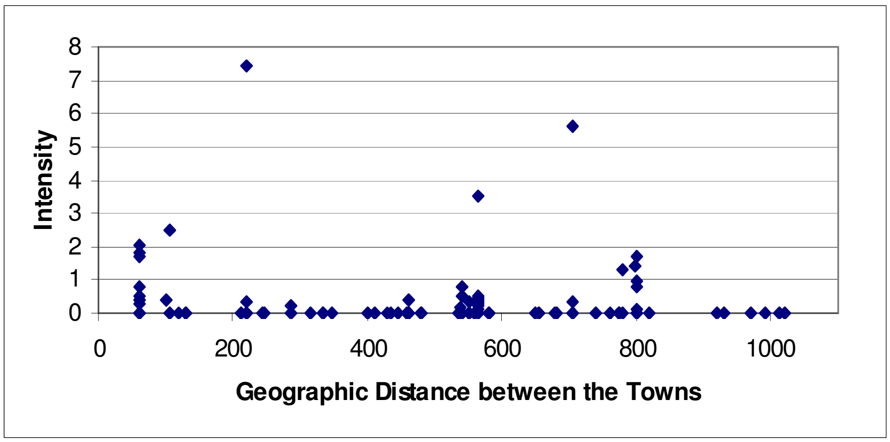


Figure 6. *Intensity of collaboration between towns versus the distance between them*

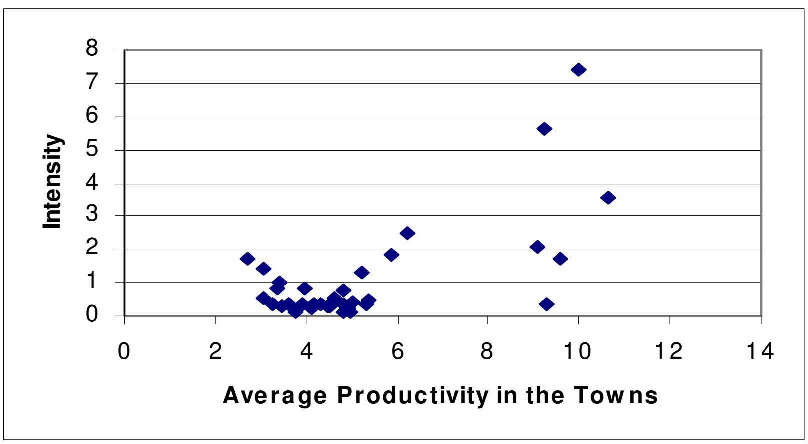


Figure 7. *Intensity of collaboration between towns versus their average productivity*

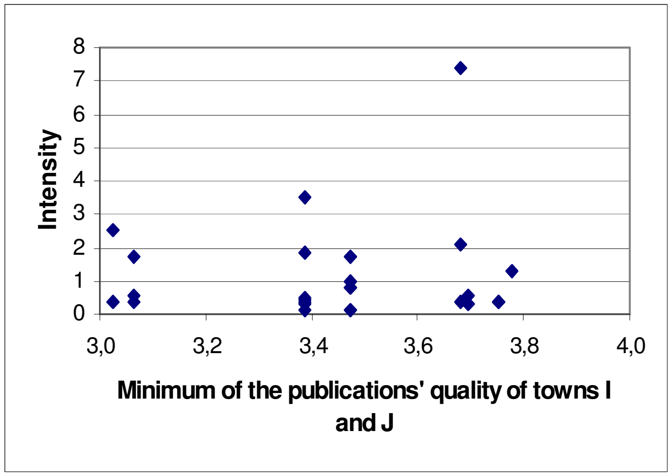


Figure 8. *Intensity of collaboration between towns versus the minimum of their publications' quality*

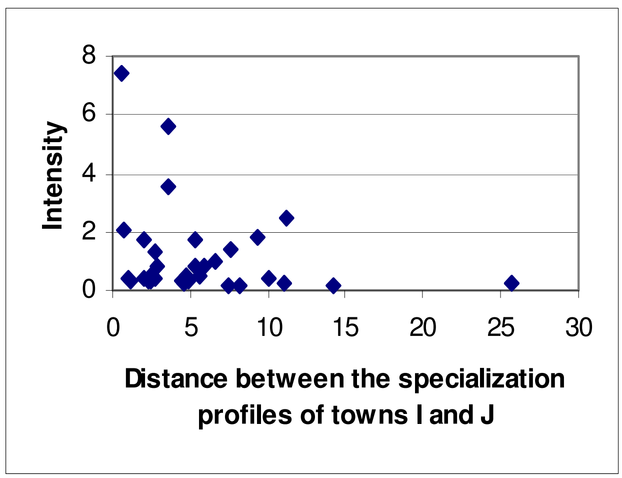


Figure 9. *Intensity of collaboration between towns versus the distance between their specialization profiles*

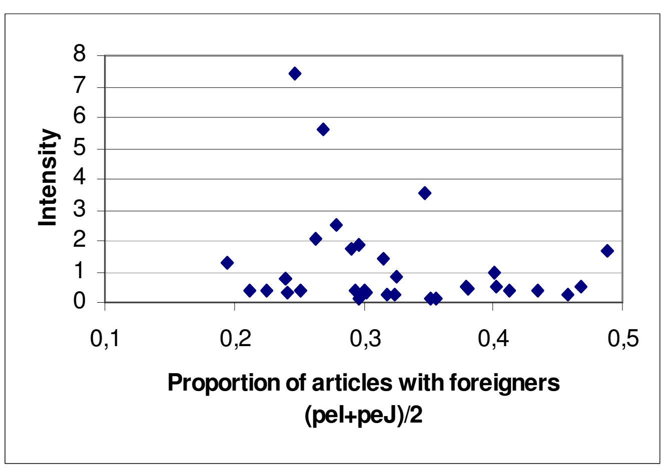


Figure 10. *Intensity of collaboration between towns versus the average proportion of their articles co-authored by foreign scientists*

Appendix 1 ***Matrix of intensities of laboratory collaboration between French towns***

	Bagneux	Poitiers	Gif sur Yvette	Grenoble	Marseille	Meudon	Mont-pellier	Orléans	Orsay	Palaiseau	Paris	Saint Martin d'Hères	Stras-bourg	Talence	Toulouse	Ville-neuve d'Ascq	Villeur-banne
Bagneux	62.96			3.53			1.73		0.37		2.57				5.62	7.4	
Poitiers		25.05															
Gif sur Yvette			4.07	0.53					0.38	1.70							
Grenoble				5.42	0.24	0.24	0.49		0.44	0.28	0.13	1.84	0.36		0.15		2.50
Marseille					9.88								1.41				
Meudon						33.33					0.80						
Montpellier							11.00		0.81	0.99	0.13	0.79			0.38		
Orléans								6.91									
Orsay									3.61	0.52	0.29	0.33			0.35		0.37
Palaiseau										4.36							
Paris											3.03	0.38				0.36	
Saint Martin d'Hères												15.39				1.29	
Strasbourg													35.16				
Talence														9.88			
Toulouse															9.63		
Villeneuve d'Ascq																38.52	
Villeur-banne																	43.21

Appendix 2
Intensity in the networks