

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2003-49

**Mesure de l'Intensité de Collaboration
dans la Recherche Scientifique
et Evaluation du Rôle de la Distance
Géographique**

L. TURNER¹
J. MAIRESSE²

Les documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.

Working papers do not reflect the position of INSEE but only the views of the authors.

¹ EUREQua, Université Paris 1 Panthéon-Sorbonne, Maison des Sciences Economiques, 106-112, Boulevard de l'Hôpital, 75013 Paris, Mail : turner@univ-paris1.fr et CREST-LEI.

² CREST-INSEE, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, et EHESS. Mail : mairesse@ensae.fr

Mesure de l'Intensité de Collaboration dans la Recherche Scientifique et Evaluation du Rôle de la Distance Géographique

Laure Turner^a et Jacques Mairesse^b

Résumé – Pour étudier les relations de collaboration entre chercheurs, nous proposons une mesure simple d'intensité de ces relations, qui a l'avantage de bien s'interpréter en termes de probabilité relative et de se prêter commodément à une agrégation au niveau du laboratoire de recherche. Nous l'appliquons pour caractériser les relations de collaboration, définies en termes de co-publication, entre les physiciens du Centre National de la Recherche Scientifique (CNRS) dans le domaine de la matière condensée pour la période 1992-1997, et pour analyser dans quelle mesure leur intensité de collaboration dépend de la distance géographique entre laboratoires. Nous trouvons ainsi que l'intensité de co-publication entre physiciens du CNRS travaillant dans un même laboratoire est en moyenne de l'ordre 40 fois plus élevée que s'ils appartiennent à des laboratoires différents mais situés dans la même ville, et de l'ordre de 100 plus élevée que s'ils sont dans des laboratoires dans des villes différentes. Toutefois l'existence de liens de co-publication et l'intensité de co-publication ne paraissent pas influencés, ou très faiblement, par la distance géographique entre villes pour les chercheurs travaillant dans des villes différentes.

Mots-clefs : économie de la science, collaboration, réseau. **Code JEL** : D29, O39

A Measure of Intensity of Collaboration in Scientific Research and an Analysis of Impact of Geographical Distance

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 it 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 analyse if their intensity of collaboration depends on the geographical distance between laboratories. 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.

Keywords: economics of science, collaboration, network. **JEL Classification**: D29, O39.

^a EUREQua, Université Paris I Panthéon-Sorbonne, Maison des Sciences Economiques, 106-112 Bd de l'Hôpital, 75013 Paris, e-mail :turner@univ-paris1.fr ; et CREST-LEI.

^b CREST-INSEE, 15 Bd Gabriel Péri, 92242 Malakoff, e-mail :mairesse@ensae.fr ; et EHESS.

Nous remercions Michèle Crance, Serge Bauin et les membres de l'Unité des Indicateurs de la Politique Scientifique (UNIPS, CNRS), pour leur aide lors de la constitution de la base de données, ainsi que pour leurs remarques Anne Crubellier et Claudine Hermann (chercheurs en Physique de la matière condensée), Dominique Foray et les participants à la Conférence en l'honneur de Paul David (Turin, 2000).

I. Introduction

Assurer et accroître l'efficacité du système scientifique est devenu une composante essentielle des politiques de la recherche et de l'innovation dans les économies contemporaines fondées sur la connaissance. Un champ très important d'analyse et d'investigation s'est ainsi ouvert, dont l'enjeu est d'éclairer la place de la science dans la dynamique économique et la croissance et le rôle des institutions scientifiques dans la production et la diffusion du savoir. L'économie de la science a notamment pour vocation d'étudier des questions aussi variées que les configurations institutionnelles des systèmes scientifiques, le marché du travail des chercheurs, leurs rémunérations, carrières et incitations spécifiques, l'allocation des fonds publics dans la recherche, ou encore les politiques scientifiques ; elle contribue ainsi à la réflexion sur une organisation performante de la science (Dasgupta et David, 1994 ; Gibbons and al., 1994 ; Diamond, 1996 ; Stephan, 1996 ; Callon et Foray, 1997 ; Shi, 2001).

Le présent travail s'inscrit dans la littérature en économie de la science sur les externalités de connaissances. Cette littérature concerne jusqu'à présent principalement les externalités géographiques qui favorisent localement l'émergence de savoir nouveau. Elle étudie plus particulièrement l'existence de ces externalités au sein de l'industrie ou entre la recherche publique et l'industrie, à partir des données de citations de brevets¹. Notre travail correspond quant à lui davantage à une étude des externalités de connaissances à l'intérieur même de l'institution scientifique. On s'intéresse aux facteurs explicatifs de l'existence des liens de collaboration dans la recherche publique et de l'intensité de ces liens. Audretsch et Stephan (1996) ont apporté une contribution qui se rapproche de la nôtre, néanmoins encore dans le cadre de l'étude des liens entre recherche publique et industrie. Les auteurs montrent, à partir de données concernant la position occupée par des scientifiques universitaires dans des entreprises de biotechnologie américaines, que les liens entre ces entreprises et les chercheurs travaillant à proximité sont d'autant plus probables que leur réputation académique est élevée, que leur réseau de collaboration reste géographiquement peu étendu, et qu'ils sont déjà impliqués dans le transfert de connaissances vers l'entreprise (ayant participé à sa création ou étant membre de son Conseil scientifique).

Nous proposons dans cet article une mesure d'intensité de la collaboration entre chercheurs qui a l'avantage de s'interpréter assez intuitivement et de s'agréger simplement à l'échelle du laboratoire, ou à celle du groupe de laboratoires dans une ville. Nous illustrons ici l'utilisation de cette mesure en considérant, parmi les nombreux facteurs susceptibles d'expliquer la formation et l'intensité de collaboration entre laboratoires de recherche, quelle

¹ On peut mentionner trois de ces études. Jaffe (1989) met en évidence aux Etats-Unis une relation étroite entre le nombre de brevets et l'importance de la recherche universitaire au niveau étatique, qu'il interprète comme le signe de l'existence d'externalités géographiques. Jaffe, Trajtenberg et Henderson (1993) rendent compte de la localisation des externalités de connaissances à partir de données de citation de brevet. Les auteurs montrent que brevets citants et brevets cités appartiennent avec une forte probabilité à la même région géographique. Jaffe et Trajtenberg (1998), toujours à partir des données de citations de brevet, font un constat similaire d'une localisation nationale des flux de connaissances : les brevets dont les inventeurs résident dans le même pays ont entre 30% et 80% de chances en plus de se citer mutuellement que si leurs inventeurs habitent des pays différents.

peut être l'importance de la distance géographique². La collaboration scientifique paraît-elle fortement limitée ou non par l'éloignement géographique des chercheurs et de leurs laboratoires ?

Dans cette étude, nous repérons les liens de collaboration à partir des données de co-publication relatives à un échantillon de près de 500 physiciens français du Centre National de la Recherche Scientifique (CNRS) sur la période 1992-1997. Nous repérons et examinons ces liens au niveau des laboratoires de recherche et des villes où ils se trouvent souvent regroupés, mais nous aurions pu le faire, dans une perspective d'analyse fine des interactions individuelles, au niveau même des chercheurs.

Le plan de l'article est le suivant. Nous précisons d'abord le champ de l'étude, la nature des données et présentons les caractéristiques générales de collaboration. Nous définissons ensuite notre mesure d'intensité de collaboration et donnons un exemple détaillé de son calcul. Enfin nous présentons les résultats obtenus à partir de cette mesure et commentons ses corrélations avec la distance géographique. Nous considérons surtout les résultats trouvés à l'échelle des villes, car ils sont très comparables à ceux trouvés à l'échelle des laboratoires et nous pouvons les présenter en plus grand détail.³

II. Champ de l'étude et caractéristiques générales de la collaboration

2.1 *Le champ de l'étude : la collaboration entre chercheurs du CNRS en physique de la matière condensée*

Nous étudions les réseaux de collaboration de 493 chercheurs en Physique de la matière condensée du CNRS sur la période 1992-1997. Ces scientifiques appartiennent à la section de la matière condensée, ils sont nés entre 1936 et 1960 et travaillaient encore au CNRS en 1997. Le domaine de la matière condensée a été choisi a priori pour deux raisons principales. D'une part, il présente des caractéristiques favorables pour notre étude : ses recherches se classent dans la science fondamentale pure, les journaux de forte réputation sont clairement identifiés, la taille du domaine couvert est bien définie et les chercheurs sont peu mobiles. D'autre part, la matière condensée est un domaine en pleine expansion, mis à l'honneur avec l'attribution du Prix Nobel de Physique à Pierre-Gilles de Gennes en 1991, qui représente actuellement près de la moitié de la Physique académique française. La physique de la matière condensée étudie, à des échelles diverses (atome, molécules, colloïdes, grains ou cellules), tous les états de la matière entre les liquides et les solides où les molécules sont relativement proches, et ce à partir d'un héritage de traditions expérimentales (cristallographie, diffusion de neutrons, d'électrons, imagerie par résonance magnétique

² Les études bibliométriques antérieures ont montré que la réputation, la popularité et visibilité des chercheurs, le besoin d'instruments spécifiques de recherche, la spécialisation thématique et la proximité géographique étaient favorables à la collaboration dans la recherche scientifique (voir Katz pour une revue de cette littérature, 1994).

³ Pour une présentation plus large et complète des résultats, et notamment une extension de l'analyse à d'autres facteurs de l'intensité de collaboration que la distance géographique, voir Mairesse et Turner (2001), et Turner (2003).

nucléaire, microscopies, ...) et théoriques (physique statistique). Elle est aussi amenée de plus en plus à développer des liens avec l'industrie, autour des matériaux de l'électronique, granulaires, des plastiques, des gels alimentaires ou cosmétiques, ...

L'échantillon de 493 physiciens que nous considérons dans cette étude représente une grande part du nombre total des chercheurs CNRS de cette discipline, puisqu'en 1996, la physique des milieux denses et des matériaux du CNRS comprenait 654 chercheurs⁴. Notre échantillon concerne ainsi pratiquement tous les physiciens en ce domaine, nés entre 1936 et 1960, et travaillant encore au CNRS en 1997. Notons qu'il a été possible suivre le changement de nom des femmes qui ont pris au cours de la période le nom de leur mari, de sorte que leur profil de publications est complet.

Il existe un avantage important à se limiter aux chercheurs du CNRS. L'appartenance des chercheurs étudiés à la même institution, le CNRS, crée une proximité organisationnelle, marquée par le partage des connaissances communes et des règles implicites ou explicites de l'organisation, qui favorise l'échange et la coordination (Rallet et Torre, 2001, Foray, 2000). Etant membre de la même institution et de la même communauté scientifique, les chercheurs disposent d'un terrain qui facilite la coopération informelle, à savoir la coopération qui n'implique pas, en amont, un temps de définition des règles de coordination. Le fait de cette proximité organisationnelle permet a contrario de mieux saisir les effets de la distance proprement géographique sur la collaboration.

L'indicateur de la collaboration que nous utilisons dans cette étude est la co-publication. Elle apparaît être un bon indice de la collaboration même si elle n'en est pas une mesure exhaustive, au sens où la collaboration peut conduire à d'autres finalités que la publication. Les données de co-publication utilisées dans notre travail sont extraites du *Science Citation Index*⁵ (SCI). Entre 1992 et 1997, les 493 chercheurs publient 7784 articles dont 7532 en collaboration (97% !). De fait, la collaboration apparaît comme le principal mode de publication. Seulement 132 chercheurs (moins d'un tiers) ont écrit des articles seuls, en plus de leurs articles en collaboration, et seulement pour un total de 252 articles sur 6 ans (un article sur trois ans).

Dans la mesure où nous souhaitons qualifier l'intensité des relations de collaboration, il nous est paru approprié de considérer qu'un article a d'autant plus de poids qu'il associe de chercheurs⁶. Nous étudions donc les réseaux «liens par liens», c'est-à-dire encore couples d'auteurs par couples d'auteurs. Un article figure ainsi dans la base d'étude que nous avons construite autant de fois qu'il fait intervenir de couples différents de chercheurs.

⁴ Le CNRS et l'Enseignement Supérieur (et plus marginalement l'INRETS) sont les seules institutions publiques de recherche dans ce domaine en France. En 1996, les physiciens de la matière condensée dans l'Enseignement Supérieur étaient au nombre de 1475, et ils étaient 16 à l'INRETS (Barré, Crance, Sigogneau, 1999).

⁵ Le *Science Citation Index* (SCI) est produit par l'*Institute for Scientific Information* (ISI). Il s'agit d'une base américaine qui couvre tous les champs disciplinaires scientifiques. Elle constituée à partir du dépouillement de plus de 3200 périodiques internationaux les plus cités. La qualité des données est remarquable et, en particulier, la couverture des publications des unités du CNRS est satisfaisante (UNIPS, 1999). Les chercheurs écrivent la presque totalité de leurs articles en anglais (95%), et les publications anglaises sont parfaitement couvertes par le SCI.

⁶ Une autre solution aurait pu être de fractionner l'article par le nombre de couples, afin de ne compter qu'une seule fois l'article. Cette solution n'aurait pas posé de problème particulier, mais nous a paru moins satisfaisante pour l'étude menée.

Nous avons choisi par ailleurs de mener l'étude au niveau du laboratoire de recherche, et même au niveau encore plus agrégé du groupe de laboratoires situés dans une même ville. Cela signifie que nous opérons par rapport au niveau élémentaire des collaborations interindividuelles un changement d'échelle tel que les réseaux de collaboration considérés sont des réseaux de laboratoires et de villes. Lorsque deux chercheurs qui collaborent appartiennent à des entités différentes, nous retiendrons que le couple formé de leurs entités collabore, et on parlera de collaboration inter-laboratoire ou inter-ville. De même, dans le cas où la collaboration entre les chercheurs a lieu au sein de la même entité, on parlera de collaboration intra entité, soit intra-laboratoire ou intra-ville. Mener l'analyse à l'échelle agrégée des laboratoires et des villes a l'avantage de permettre d'illustrer plus commodément l'utilisation de la mesure d'intensité de collaboration proposée. Surtout cela permet de caractériser l'influence sur la collaboration du fait de travailler dans un même laboratoire ou une même ville (et donc d'une proximité et de relations de face à face aisées).

2.2 *Deux configurations de collaboration et autres caractéristiques*

Les articles peuvent associer des chercheurs appartenant à la même institution ou non, au même pays ou non. Dans notre étude, nous utilisons une typologie simplifiée de l'affiliation des auteurs. Nous distinguons les 493 chercheurs du CNRS qui collaborent, nommés plus brièvement dans la suite « chercheurs CNRS », des autres scientifiques que l'on nommera « extérieurs », qu'ils soient universitaires ou d'une autre institution, et français ou étrangers, et nous identifions ainsi deux formes ou « modes » de collaboration. Le premier mode peut associer au moins deux chercheurs CNRS et éventuellement des « extérieurs », tandis que le second associe au plus un chercheur CNRS à des extérieurs.

Comme on peut le voir sur le *schéma de l'annexe 1*, au premier mode de collaboration (au moins deux chercheurs CNRS et des extérieurs) correspond 1823 articles, et au second (au plus un chercheur CNRS et des extérieurs) correspond 5709 articles. Les 493 physiciens qui collaborent le font le plus souvent selon ces deux modes. Néanmoins, 38 physiciens ne collaborent qu'avec d'autres chercheurs CNRS de notre liste, et jamais avec des extérieurs. Et à l'inverse, 69 chercheurs ne collaborent qu'avec des extérieurs, et jamais avec un autre chercheur CNRS de notre liste. Les 1823 articles correspondant au premier mode de collaboration – « groupe 1 » - associent donc 424 des chercheurs ($= 493 - 69$), tandis que les 5709 articles du second mode – « groupe 2 » - impliquent 455 chercheurs ($= 493 - 38$).

La collaboration implique dans les deux cas, comme le montre en détail le *tableau 1*, un grand nombre de chercheurs extérieurs. Seuls 82 articles sont écrits par des chercheurs CNRS uniquement, sans coauteurs extérieurs. Le nombre moyen d'auteurs par article est de 5.9 pour le groupe 1 et de 4.9 pour le groupe 2, et le nombre moyen de chercheurs extérieurs par article s'élève pour ces deux groupes à 3.7 et à 3.9 respectivement. A l'inverse, les articles du groupe 1 impliquent rarement plus de deux chercheurs CNRS : le nombre moyen d'auteurs CNRS par article est à peine supérieur à 2 et seuls 20% de ces articles associent en fait plus de deux chercheurs CNRS.

[tableau 1]

Notre travail concerne ici les 1823 articles du groupe 1, à savoir ceux qui ont pour auteurs au moins deux chercheurs CNRS. Trois raisons déterminent ce choix⁷.

La première raison est que la densité des relations de co-publication entre chercheurs doit être suffisamment notable pour permettre l'étude des réseaux de collaboration. De ce point de vue, le nombre de relations de collaboration par couple correspondant aux articles du groupe 1 est plus important que le nombre de relations correspondant au groupe 2. En effet, dans le groupe 1, les 1823 articles sont le fait de 424 chercheurs CNRS (et de 3500 extérieurs environ), et correspondent à 880 couples différents associant au moins 2 CNRS. En revanche, pour le groupe 2, les 5709 articles font intervenir 455 chercheurs CNRS et près de 10 000 extérieurs, et correspondent à 17 500 couples différents contenant au plus un CNRS. Ainsi, le nombre moyen d'articles par couple sur six ans est de 2.1 pour le groupe 1, contre seulement 0.33 pour le groupe 2.

La seconde raison est pratique, elle tient à ce que les chercheurs CNRS sont les seuls pour lesquels nous connaissons précisément les adresses de leurs laboratoires⁸. Une identification fiable de la localisation des extérieurs n'est en effet pas réalisable en utilisant le seul Science Citation Index à partir duquel sont construites nos données⁹. Nous ne pouvons donc pas saisir la géographie des liens de collaboration entre chercheurs CNRS et extérieurs, mais seulement entre chercheurs CNRS.

La troisième justification au choix d'un champ d'étude limité aux co-publications impliquant au moins deux chercheurs CNRS est d'ordre analytique a priori, et a déjà été soulignée. En étudiant la collaboration entre chercheurs CNRS, et en l'analysant de surcroît au niveau des laboratoires et des villes, nous faisons abstraction des effets de distance organisationnelle sur la collaboration, et nous pouvons espérer dégager ainsi les effets d'une distance proprement géographique.

Comme on peut le voir sur la carte présentée en *annexe 2*, la répartition géographique des physiciens est fortement inégale, puisque les chercheurs se concentrent essentiellement dans les laboratoires de la région grenobloise (136 chercheurs) et de l'Ile de France (204 chercheurs). A cet égard, nous avons été amenés à faire une sélection sur les villes et sur les laboratoires en fonction de leurs effectifs, afin de retenir seulement des entités suffisamment grandes pour que l'étude de l'intensité de collaboration soit pertinente. Nous imposons que les villes comportent au moins 9 chercheurs, et les laboratoires au moins 5 chercheurs. L'échantillon étudié passe ainsi de 493 à 470 chercheurs répartis dans 17 villes et 34 laboratoires. Pour la même raison, nous ne considérons pas comme collaborant significativement les couples de villes liées par moins de 6 articles et les couples de laboratoires liés par moins de 4 articles. Cette contrainte ne réduit pas le nombre de villes et

⁷ Les groupes 1 et 2 de collaboration présentent certaines caractéristiques très comparables, par exemple en termes de distribution du nombre d'articles en fonction du nombre de coauteurs extérieurs, ou de la concentration du nombre d'articles par auteur CNRS.

⁸ Grâce à une information directe communiquée par l'Unité des Indicateurs de la Politique Scientifique (UNIPS), CNRS.

⁹ Dans le Science Citation Index, la localisation des auteurs se fonde sur les indications figurant dans les articles publiés, qui sont difficiles à utiliser. Le nombre d'auteurs recensés pour un article est souvent différent du nombre d'adresses listées. Il est par exemple possible que plusieurs collaborateurs aient la même adresse, auquel cas celle-ci n'apparaît qu'une fois dans la liste des adresses. Mais on ne peut savoir à qui, parmi l'ensemble des auteurs, attribuer cette adresse lorsque la collaboration implique aussi d'autres laboratoires. Un autre exemple fréquent est celui des affiliations multiples, où un auteur indique son appartenance à plusieurs laboratoires, ce qui fait que le nombre d'adresses devient supérieur au nombre d'auteurs, avec à nouveau un problème d'attribution précise.

de laboratoires considérés, mais conduit simplement à « compter pour zéro », dans les bases d'études agrégées au niveau des villes et des laboratoires, les collaborations relatives à moins de 6 ou 4 co-publications (sur six ans) entre chercheurs CNRS des villes et laboratoires concernés. Au total, le nombre d'articles écrits en collaboration par au moins deux chercheurs CNRS (groupe 1) sous-jacent au choix de ces seuils passe de 1823 à 1634 articles¹⁰.

III. Une mesure d'intensité de collaboration

L'insertion des agents dans les réseaux est déterminée par des caractéristiques individuelles « intrinsèques », telles que l'âge, le sexe, la capacité, la stratégie, et par des variables plus structurelles, comme le nombre de relations qu'ils développent, ou leur distance géographique, De ce fait, les réseaux diffèrent par leur forme et leur fonctionnement. Si les acteurs étaient indifférenciés, et qu'ils collaboraient de manière équiprobable avec tous les autres, on s'attendrait au contraire à observer une structure uniforme de liens reliant tous les individus. Nous prenons ce cas d'homogénéité comme cas de référence. Au niveau agrégé des laboratoires et des groupes de laboratoires, ou villes, auquel on se place dans ce travail, le cas d'homogénéité correspond à la configuration dans laquelle la fréquence de collaboration de chaque entité avec toutes les autres serait la même, quel que soit leur localisation géographique et leurs attributs propres. Notre mesure d'intensité de la collaboration entre deux entités est fondée simplement sur la comparaison entre le réseau réel, tel que l'on peut le caractériser effectivement sur la base des données observées, et celui que l'on observerait en cas d'homogénéité parfaite.

3.1 Définition

Dans cette section et la suivante, on suppose pour simplifier que la collaboration associe toujours deux CNRS au plus¹¹. Le réseau étudié comporte un nombre fini d'entités, qui regroupent au total N chercheurs, pouvant former C couples de collaboration (où C est donc égal à $N(N-1)/2$, nombre des combinaisons possibles deux à deux). Si n est le nombre total d'articles issus de la collaboration entre les N chercheurs, la fréquence p du nombre de co-publications par couple pour l'ensemble du réseau est le rapport entre ce nombre total d'articles, n , et le nombre de couples possibles, C . C'est aussi celle que l'on observerait entre deux entités quelconques de ce réseau sous l'hypothèse d'homogénéité des entités qui le constituent et de leurs collaborations (voir infra).

Considérons à présent deux entités X et Y de ce réseau. N_X chercheurs travaillent dans l'entité X et N_Y chercheurs dans l'entité Y . Les nombres de couples possibles au sein de X et de Y sont notés $C_X = (N_X)(N_X-1)/2$ et $C_Y = (N_Y)(N_Y-1)/2$, tandis que le nombre de couples possibles pouvant se former entre des chercheurs de X et de Y est noté $C_{XY} = (N_X)(N_Y)/2$. De même, les nombres d'articles écrits au total en collaboration au sein de X et de Y sont notés respectivement n_X et n_Y , et le nombre total d'articles écrits entre chercheurs de X et de Y est noté n_{XY} . La fréquence observée p_{XY} de collaboration entre les deux entités X et Y est égale

¹⁰ Notons que les quelques variantes de calcul que nous avons faites montrent que le choix de seuils sensiblement différents n'affecte pas la physionomie de nos résultats.

¹¹ Ce point est discuté dans la section 3.3.

au rapport entre le total d'articles écrits par les chercheurs de X et de Y en commun et le nombre de couples de chercheurs possibles entre ces deux entités. De la même manière, la fréquence observée de collaboration au sein d'une même entité X ou Y, notée p_X ou p_Y , est le rapport entre le total d'articles écrits par les chercheurs de X ou Y entre eux et le nombre de couples de chercheurs possibles dans l'entité X ou Y. On a donc :

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

On définit alors les intensités de collaboration *intra*-entités et *inter*-entités pour X et Y comme le rapport des fréquences observées à la fréquence p obtenue pour l'ensemble du réseau :

$$i_X = \frac{p_X}{p}, \quad i_Y = \frac{p_Y}{p}, \quad \text{et} \quad i_{XY} = \frac{p_{XY}}{p}$$

En cas de parfaite d'homogénéité du réseau, on doit avoir $p_X = p_Y = p_{XY}$ pour tout X et Y, d'où il apparaît simplement (par sommation des numérateurs et des dénominateurs pour toutes les entités) que $p_X = p_Y = p_{XY} = p$, ou encore en termes de la mesure d'intensité $i_X = i_Y = i_{XY} = 1$. En cas d'homogénéité, la fréquence pour l'ensemble du réseau, p , est celle que l'on s'attend à trouver pour X et Y, aussi bien en *intra* qu'en *inter*. La raison en est évidente puisqu'on agrège en ce cas au niveau des entités des probabilités de collaboration entre individus dont on suppose qu'elles sont égales dans tout le réseau.

Ainsi, en cas d'homogénéité du réseau, les intensités *intra* et *inter* de collaboration valent 1. Dans le cas contraire, qui se présente pour un réseau réel, les intensités sont influencées par différents facteurs –par exemple la distance que l'on étudie ici- de sorte qu'elles s'expriment comme un rapport de fréquences observées de collaboration qui est généralement différent de 1.

Notons que nous pouvons représenter la « structure d'intensité » d'un réseau de E entités par une matrice symétrique $E \times E$ à coefficients positifs ou nuls, dont les termes diagonaux sont les intensités *intra entités*, et les autres termes les intensités *inter entités*. La matrice d'adjacence du réseau utilisée en théorie des graphes qui indique seulement l'existence des liens de collaboration entre entités s'en déduit immédiatement : les coefficients nuls des deux matrices sont les mêmes, et aux coefficients positifs de la première correspond les coefficients unités de la seconde.

3.2 Propriété d'agrégation

L'intensité telle que nous venons de la définir présente l'avantage de s'agréger facilement. Pour le voir, prenons deux entités X et Y (laboratoires), composant une entité V (ville) plus vaste. Le nombre total d'articles écrits en collaboration dans V est la somme des articles écrits en collaboration au sein de X et Y, et en collaboration entre chercheurs de X et Y. De même, le nombre de couples possibles de chercheurs de V est la somme des couples possibles de chercheurs de X et Y, et entre X et Y. On a ainsi:

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

Cette formule peut s'écrire également de la manière suivante :

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

ou encore :

$$p_V = p_X * \frac{C_X}{C_V} + p_Y * \frac{C_Y}{C_V} + p_{XY} * \frac{C_{XY}}{C_V}$$

Nous avons ainsi une formule d'agrégation, qui se généralise à un réseau E d'entités I (laboratoires). En divisant la formule par p, nous obtenons l'intensité de collaboration d'un réseau comme la somme pondérée des intensités intra et inter entités. Les pondérations représentent le poids de chaque entité dans le réseau en terme de couples possibles de collaboration :

$$i_E = \sum_I i_I w_I + \sum_{I,J \neq I} i_{IJ} w_{IJ}$$

$$\text{où } w_I = \frac{C_I}{C_E} \quad \text{et } w_{IJ} = \frac{C_{IJ}}{C_E} \quad \text{avec } \sum_{I,J} w_{IJ} = 1$$

3.3 Remarques sur la pondération

Jusqu'à présent, nous avons supposé que les articles étaient écrits par deux chercheurs au plus. En réalité, ils peuvent être écrits aussi par des triplets, ou des quadruplés, ..., de chercheurs. Or, comme nous l'avons indiqué (section 2.1), dans la mesure où nous intéressons à la formation et à l'intensité des liens de collaboration, nous avons choisi comme unité d'analyse les couples de collaboration, et construit notre base d'étude de telle sorte qu'un article y est répété autant de fois qu'il a de couples d'auteurs différents. Autrement dit, nous dissociions les triplets, les quadruplés, ..., en paires, de sorte que le nombre d'articles écrits en collaboration considéré pour le calcul de l'intensité est un nombre pondéré par le nombre de couples qui contribuent à sa publication.

On aurait pu envisager bien sûr de ne pas faire un tel calcul pondéré. On aurait pu compter, par exemple, trois fois un tiers de lien, pour un article ayant trois auteurs CNRS (soit trois couples), autrement dit fractionner un tel article en proportion du nombre de couples d'auteurs CNRS, afin de ne compter finalement qu'une seule fois l'article. Dans la mesure où la proportion d'articles écrits en collaboration par plus de deux chercheurs CNRS est faible, les deux types de calcul conduisent dans le cas présent à des résultats peu différents. Remarquons également que l'on fait l'hypothèse implicite que deux liens de collaboration établis par deux couples différents ont la même importance que deux liens de collaboration

établis par le même couple. Autrement dit, une collaboration peut paraître intense entre deux entités, ou au sein d'une entité, alors qu'à la limite elle ne serait due en réalité qu'à un seul couple d'auteurs¹².

3.4 Calcul pratique : un exemple

Prenons un exemple précis, celui de Marseille, pour décrire le calcul de l'intensité de collaboration, en faisant référence au *tableau 2* qui présente également le résultat de ce calcul pour les autres villes étudiées. Marseille est une ville qui rassemble 18 physiciens de notre liste (colonne 1). Elle développe 34 collaborations en son sein (colonne 3), 18 liens de collaboration avec d'autres villes (colonne 4), dont 10 liens avec Grenoble, et 8 avec Strasbourg, qui sont ses deux *partenaires effectifs*. Elle a aussi des relations avec Poitiers, Gif sur Yvette, Orsay, Toulouse, Villeurbanne, mais qui ne sont pas prises en compte car elles sont en nombre inférieur à 6.

Le nombre de couples possibles de chercheurs travaillant à Marseille est $18 \times 17 / 2$, soit 153. La fréquence observée du nombre de collaborations par couple dans Marseille vaut donc $34 / 153$, soit 0.2222. Sachant que les effectifs de chercheurs à Grenoble et Strasbourg sont respectivement de 105 et 14, le nombre de couples possibles de chercheurs associant Marseille et Grenoble est de 1890 (105×9), et celui associant Marseille et Strasbourg de 252 (14×18). La fréquence observée du nombre de collaborations par couple entre Marseille et Grenoble vaut donc 0.0053 ($10 / 1890$), et celle entre Marseille et Strasbourg vaut 0.0317 ($8 / 252$).

Afin de donner les intensités de collaboration, il faut calculer la fréquence de collaboration par couple pour l'univers des 17 villes. Il s'agit du rapport entre le total d'articles pondérés, qui vaut 2480, et le nombre de couples possibles que peuvent former les 470 chercheurs de l'univers, soit 110215 couples ($470 \times 469 / 2$). On a donc $p = 0.0225$. En cas d'homogénéité, cette fréquence est celle que l'on aurait trouvée pour la collaboration intra comme inter de Marseille. En réalité, la fréquence observée en intra est plus élevée (0.2222), et au contraire, la fréquence de collaboration entre Marseille et Grenoble est plus faible (0.0053). La fréquence de collaboration entre Marseille et Strasbourg est plus proche de la référence (0.0317). Les intensités intra et inter sont le rapport entre les fréquences observées de collaboration et p . L'intensité intra pour la ville de Marseille vaut donc 9.88 (colonne 5). L'intensité entre Marseille et Grenoble est de 0.24, et celle entre Marseille et Strasbourg vaut 1.41, soit une intensité moyenne de collaboration de Marseille avec ses *partenaires effectifs* de 0.8 (colonne 7). La moyenne de l'intensité de collaboration de Marseille avec l'ensemble des 16 autres villes vaut 0.1 (colonne 6).

[tableau 2]

¹² La littérature économique des réseaux considère aussi une approche plus complexe, qui tient compte du nombre de liens indirects (et non seulement directs) entre entités. Plus généralement encore, les décisions individuelles d'intégrer un réseau sont fondées sur l'évaluation rationnelle par les agents de la valeur économique produite par le réseau, valeur qui dépend de sa structure. Par exemple, dans Jackson et Wolinsky (1996), la valeur produite par le réseau pour un agent est la somme des valeurs de tous ses liens directs et indirects, qui sont fonction de sa distance aux autres agents, cette somme étant diminuée par un coût proportionnel au nombre de liens.

IV. Résultats

Dans cette partie nous commentons d'abord les résultats du calcul de l'intensité de collaboration, principalement à l'échelle des villes (section 4.1), puis les corrélations trouvées entre la distance géographique et l'existence et l'intensité des liens de collaboration à l'échelle des villes et celle aussi des laboratoires (section 4.2).

4.1 Intensité de collaboration entre villes et laboratoires

Les résultats détaillés du calcul de l'intensité de collaboration dans les réseaux des villes françaises se présentent sous la forme de la matrice symétrique donnée en *annexe 3* (où les termes diagonaux sont les intensités de collaboration intra-ville, et les termes non diagonaux les intensités inter-ville, nuls en cas d'absence de collaboration). Ces résultats sont également illustrés par la *carte de France (annexe 2)*. Ils montrent clairement que les différences d'intensité inter ville sont très fortes, pouvant aller de l'existence de liens avec la plupart des villes pour la région grenobloise et la région parisienne jusqu'à une absence de lien pour trois villes : Orléans, Poitiers, et Talence. Le *tableau 2*, que nous commentons plus en détail maintenant, compare par ville les intensités moyennes inter-ville et les intensités intra-ville et apporte des précisions complémentaires.

La deuxième colonne du *tableau 2* dénombre pour chacune des 17 villes le nombre de ses partenaires dans la collaboration ; il est de 4 en moyenne, ce qui peut paraître plutôt faible¹³. Neuf des 17 villes ont moins de quatre partenaires : outre les villes déjà citées qui n'ont pas de liens de collaboration extérieure, il s'agit de Gif-sur-Yvette, Meudon, Marseille, Strasbourg, Villeneuve-d'Ascq et Villeurbanne ; pourtant trois d'entre elles ont des effectifs relativement élevés, supérieurs à 14. Les villes qui ont le plus de liens de collaboration sont Grenoble, Orsay et Paris. Ce sont celles aussi qui rassemblent le plus de chercheurs. Elles fonctionnent chacune comme un pôle en suscitant la collaboration d'une forte proportion des autres villes : 12 d'entre elles pour Grenoble (soit 70% des villes), 9 pour Orsay (53%) et 7 pour Paris (41%). Une première différence existe donc entre les villes selon le nombre de leurs partenaires. La collaboration avec de nombreux partenaires est le fait de peu de villes. Il est remarquable ainsi que sur les 136 couples de collaboration possibles pouvant se former entre les villes, seuls 34 s'établissent. Pour les laboratoires la proportion est encore plus faible : sur les 561 couples de collaboration possibles de laboratoires, seulement 56 couples se forment.

L'intensité de collaboration intra-ville figure dans la colonne 5 du *tableau 2*. Elle est élevée, toujours supérieure à 1, et en moyenne de 18.9. Le nombre de collaboration intra-ville est donc en moyenne près de 19 fois supérieur au chiffre que l'on pouvait attendre en cas d'homogénéité des entités, ce qui montre l'importance de la collaboration à l'intérieur du périmètre de la ville. Notons aussi que Poitiers, Meudon, Strasbourg, Villeneuve-d'Ascq et Villeurbanne, qui ont peu de partenaires, ont en revanche des niveaux d'intensité intra de collaboration qui sont 25 fois supérieurs aux niveaux auxquels on pouvait s'attendre en cas

¹³ Il convient de rappeler que nous caractérisons l'existence d'un lien de collaboration entre deux villes si les chercheurs CNRS travaillant dans ces villes en physique des matières condensées ont écrit ensemble au moins 6 articles sur la période d'études 1992-1997 (soit en moyenne au moins un article par an). Ce nombre de 4 partenaires dans la collaboration en moyenne peut néanmoins paraître relativement faible pour des villes regroupant au moins 9 chercheurs CNRS en ce domaine, mais nous n'avons pas de point de comparaison avec d'autres études.

d'homogénéité. En revanche, les pôles liés à de nombreux partenaires comme Grenoble, Orsay et Paris, appartiennent au groupe des villes qui présentent les intensités intra-ville les plus faibles, soit 5.4, 3.6 et 3.0 respectivement. Cette différence s'explique néanmoins par le fait que Grenoble, Orsay et Paris comprennent plusieurs laboratoires qui collaborent entre eux, alors que les premières villes mentionnées ont un unique laboratoire. De fait, les intensités moyennes de collaboration intra à l'échelle des laboratoires sont de 19.4, 33.4 et 34.5 pour ces trois villes respectivement, et donc très comparables à celles des autres villes (voir *tableau 3*).

L'intensité moyenne par ville de collaboration inter-ville est présentée dans les colonnes 6 et 7 du *tableau 2*. Comme on l'a dit, sur les 136 couples possibles de villes, 34 couples collaborent effectivement. Les colonnes 6 et 7 indiquent pour chaque ville la moyenne de l'intensité de collaboration inter-ville, calculée respectivement sur l'ensemble des 16 autres villes et sur les seuls partenaires effectifs. La moyenne calculée sur les partenaires effectifs vaut 1.0 en moyenne, tandis que celle calculée sur l'ensemble des villes n'est que de 0.3 en moyenne, et est toujours, sauf pour Bagnex, inférieure à 1. La moyenne sur l'ensemble des villes tient compte, par définition, des nombreux liens de collaboration possibles qui ne se développent pas. Orsay et Paris, qui devraient former de nombreux liens étant donnés leurs effectifs, présentent ainsi des moyennes d'intensités de collaboration inter-ville sur l'ensemble des villes parmi les plus faibles (0.2 et 0.3), alors que des villes plus petites comme Bagnex et Villeneuve-d'Ascq ont des moyennes d'intensités plus élevées (1.3 et 0.6). Néanmoins, Grenoble, ville qui a le plus de partenaires, a aussi une moyenne d'intensité inter-ville de même ordre (0.7).

Dans tous les cas, la collaboration inter-ville est beaucoup moins intense que la collaboration intra-ville, indiquant clairement le rôle positif très fort de la proximité immédiate sur l'existence et l'intensité des liens de collaboration. Cette observation est même renforcée si l'on considère les intensités de collaboration à l'échelle des laboratoires présentées dans le *tableau 3*. En effet, en comparant l'intensité de collaboration intra-laboratoire avec l'intensité de collaboration entre laboratoires partenaires situés au sein d'une même ville, et avec l'intensité de collaboration entre laboratoires partenaires situés dans des villes différentes, on obtient le classement suivant : l'intensité de collaboration intra-laboratoire est environ 40 fois plus élevée en moyenne que l'intensité de collaboration inter-laboratoire mais intra-ville, elle-même deux fois et demie plus élevée en moyenne que l'intensité de collaboration inter-laboratoire et inter-ville. On peut noter cependant que Paris fait exception, avec une intensité moyenne de collaboration entre les laboratoires qui est très faible (0.2) en comparaison avec celle des autres villes possédant plusieurs laboratoires, et notamment par rapport à Grenoble pour laquelle cette intensité est de 2.7. L'examen plus précis à l'échelle des laboratoires montre que c'est moins l'intensité des liens créés qui diffère que le nombre de liens qui se forment : à Paris, 4 liens de collaboration se forment sur 15 possibles, alors que 10 se forment à Grenoble sur 15 possibles. L'explication est sans doute liée en grande partie à des différences de spécialisation des laboratoires parisiens plus marquées que pour les laboratoires grenoblois ; il n'est pas impossible aussi qu'elle traduise pour une part les difficultés très grandes de circulation à Paris, qui rendent les rencontres face-à-face moins aisées.¹⁴

¹⁴ Zucker, Darby et Armstrong, 1994, dans une étude sur les externalités de recherche entre les laboratoires publics et l'industrie, montre ainsi que le savoir à l'origine de ces externalités se transmet plus aisément lors de face-à-face entre les acteurs concernés, car il revêt souvent un caractère fortement tacite. Voir également Leamer et Stopper, 2000.

[tableau 3]

4.2 Effets de la distance géographique sur l'établissement des liens de collaboration et sur leur intensité

La comparaison des intensités de collaboration intra et inter villes et laboratoires conduit à distinguer qualitativement trois sortes de distances : la proximité immédiate (distance nulle), la distance faible à l'intérieur d'une ville qui est probablement perçue davantage comme un temps de déplacement, et la distance géographique véritable entre villes. Nous essayons maintenant de voir quels peuvent être les effets de la distance géographique sur la constitution des liens inter-ville de collaboration et, si ces liens existent, ses effets sur leur intensité. Nous nous contentons pour cela de calculer les corrélations simples correspondantes.¹⁵

Le *tableau 4* donne pour chaque ville, les distances moyennes qui les séparent de leurs partenaires effectifs. On voit notamment que ces distances moyennes peuvent varier assez fortement d'une ville à l'autre. Ainsi, Montpellier collabore avec des villes distantes en moyenne de 550 km, tandis que Gif-sur-Yvette a des liens avec des villes proches, situées en moyenne à 170 km. Si on considère la distribution complète des distances entre couples de villes, on note toutefois que parmi les cinq villes situées à moins de 300 km de leurs partenaires, quatre appartiennent à la région parisienne (laquelle regroupe en fait 6 villes sur les 17).

[tableau 4]

Le *tableau 5* présente les corrélations obtenues à l'échelle des villes et à l'échelle des laboratoires¹⁶. La première colonne donne la corrélation de la distance géographique avec une variable indicatrice de l'existence de collaboration entre les entités, et la seconde la corrélation de la distance avec l'intensité de collaboration entre entités effectivement partenaires.

[tableau 5]

A l'échelle des villes, les coefficients de corrélation sont négatifs mais assez faiblement, et sans être statistiquement significatifs. Les chances de collaborer entre chercheurs de villes différentes ne seraient donc pas sensiblement moins grandes, ni l'intensité de leur collaboration moins élevée, si ces villes sont éloignées que lorsqu'elles sont proches.

A l'échelle des laboratoires, on observe cependant une corrélation négative (-0.18), fortement statistiquement significative, entre la distance géographique et la constitution de liens de collaboration. Cette observation tient cependant largement au fait que l'intensité de collaboration entre laboratoires au sein des villes est particulièrement élevée. Elle s'explique

¹⁵ On peut trouver une analyse plus complète, mais dont les conclusions ne sont pas différentes, dans Mairesse et Turner, 2001.

¹⁶ Pour le calcul des corrélations à l'échelle des laboratoires, nous avons supposé que les laboratoires situés dans une même ville étaient séparés par une distance nulle.

donc en bonne partie par l'importance d'une très grande proximité qui rend les contacts beaucoup plus faciles entre chercheurs d'une même ville, quand bien même ils sont dans des laboratoires différents. Elle correspond aussi au fait que dans ce cas la corrélation à l'échelle des 34 laboratoires est estimée plus précisément car elle est calculée sur un échantillon quatre fois plus grand que celle calculée à l'échelle des 17 villes. En ce qui concerne l'intensité de collaboration entre laboratoires, il apparaît en revanche que la distance géographique ne joue pas de façon statistiquement significative.

Enfin, si l'on considère uniquement la collaboration entre laboratoires situés dans des villes différentes (collaboration inter-laboratoire inter-ville), on observe que la corrélation entre la distance et l'établissement de la collaboration est encore statistiquement négative, mais deux fois plus faible (-0,09), la corrélation entre distance et intensité de collaboration restant non significative. Il semblerait donc au total que la collaboration soit défavorisée, mais légèrement seulement, par l'éloignement géographique.

5 Conclusion

Notre étude a eu pour objet de proposer un indicateur de collaboration, qui mesure simplement et de façon assez intuitive son intensité, et qui permet notamment d'analyser l'influence de différents facteurs sur l'établissement et l'intensité de collaboration. La mesure construite est utilisée pour étudier l'un de ces facteurs, la distance géographique. Nous montrons que la proximité géographique, définie par le périmètre de la ville, favorise la collaboration : la collaboration intra-ville est très élevée, et les laboratoires ont d'autant plus de chances de collaborer qu'ils sont proches (même si les laboratoires parisiens forment une exception). Au-delà de la proximité immédiate, la distance géographique semble décourager légèrement l'établissement des liens de collaboration entre les laboratoires, mais elle n'est pas un facteur déterminant pour expliquer la dispersion des intensités de collaboration.

Nos résultats suggèrent ainsi l'existence de trois sortes de distances : la proximité immédiate (distance nulle) qui favorise les échanges sous forme de face-à-face entre les chercheurs d'un même laboratoire ; la distance entre laboratoires au sein d'une même ville, qui est associée à des intensités de collaboration plus faibles que les intensités intra-laboratoire, et qui relève sans doute davantage d'un temps de déplacement ; et enfin la distance proprement géographique inter-ville. Lorsque l'on s'affranchit des relations de face-à-face et de voisinage, la distance géographique n'apparaît pas vraiment décisive dans les choix de collaboration. Les nouvelles technologies de communication ont certainement contribué dans les années récentes à beaucoup atténuer le rôle de la distance géographique en favorisant considérablement les échanges entre chercheurs (courrier électronique, envoi de documents de travail et articles, connexion à des bases communes de données ...).

La mesure d'intensité proposée peut être utilisée dans une analyse multifactorielle pour étudier d'autres déterminants de la collaboration que la distance géographique, et pour quantifier leur contribution respective à l'explication de la propension des chercheurs à collaborer. Notamment, il est possible que les collaborations soient dictées par la spécialisation des laboratoires, leur taille, leur productivité, leur qualité, leur ouverture

internationale, ...¹⁷ Cette mesure d'intensité peut aussi être utile dans des analyses à l'échelle des chercheurs individuels, par exemple pour l'étude des réseaux de recherche se formant autour des chercheurs « stars », qui contribuent fortement aux publications dans leur discipline scientifique.¹⁸

¹⁷ Voir Mairesse et Turner (2001). En particulier, l'étude de l'effet de la spécialisation des laboratoires est intéressante. La *carte de France* présentée en *annexe 2* suggère en effet l'influence de la spécialisation sur la géographie des réseaux de co-publication. Si la distance joue peu dans l'intensité des relations, il semble que ce soit aussi parce que les collaborations sont dictées pour une large part par la spécialisation des laboratoires. En particulier, les très grands équipements utilisés par les physiciens de la matière condensée, les anneaux de stockage, sont présents dans essentiellement deux laboratoires, à Orsay et à Grenoble, qui apparaissent de ce fait comme des pôles centraux dans la collaboration. Il n'existe pas de données de spécialisation « type », qui puissent être commodément utilisées dans les études. L'information sur la spécialisation des laboratoires qui peut être plus ou moins inféré sur la base de la spécialisation scientifique des publications demanderait à être appuyée par des enquêtes auprès des laboratoires.

¹⁸ Voir Turner (2003).

Bibliographie

- 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 émergente 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 : University 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 : University 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., 1994, « Toward a New Economics of Science », *Research Policy*, 23(5), 487-521.
- DIAMOND A., 1996, « The Economics of Science », *Knowledge and Policy*, 9 (2-3).
- FORAY D., 2000, « L' Economie de la Connaissance », La Découverte, Collection Repères.
- GIBBONS M. et al., 1994, « *The New Production of Knowledge* », Sage.
- JACKSON M.O., WOLINSKY A., 1996, « A Strategic Model of Social and Economic Networks », *Journal of Economic Theory*, 71, 44-74.
- 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), 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.
- MAIRESSE J., L.TURNER, 2001, « Measurement and Explanation of the Intensity of Co-Publication in Scientific Research: An Analysis at the Laboratory Level », in revision. To appear in “ *New Frontiers in the Economics of Innovation and New Technolog: Essays in honour of Paul David*”, eds C.Antonelli, D. Foray, B.Hall and E.Steinmueller, Oxford University Press.
- RALLET A., TORRE A., 2001, « Proximité géographique ou proximité organisationnelle? Une analyse spatiale des coopérations technologiques dans les réseaux localisés d’innovation », *Economie Appliquée*, 54(1), 147-171.
- SHI Y., 2001, « *The Economics of Scientific Knowledge* », Edward Elgar.
- 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.
- STEPHAN P., 1996, « The Economics of Science », *Journal of Economic Literature*, 34, 1199-1235.
- TURNER L., 2003, « La recherche publique dans la production de connaissances. Contributions en économie de la science », *Thèse de doctorat*, Université Paris I.
- UNIPS, 1999, « *Les Publications des Laboratoires du CNRS et leur Impact* », Éditions CNRS, mars.
- ZUCKER L, DARBY M., ARMSTRONG J., 1994, « Intellectual Capital and the Firm : the technology of geographically localized knowledge spillovers », *NBER Working Paper 4946*.
- ZUCKER L, DARBY M., ARMSTRONG J., 1994, « Intellectual Capital and the Firm : the technology of geographically localized knowledge spillovers », *NBER Working Paper 4946*.

Tableau 1 *Nombre d'articles écrits en collaboration pour le groupe 1 et 2 en fonction du nombre et du type des auteurs (entre parenthèses figure l'effectif de chercheurs CNRS concernés)*

Nombre d'articles écrits en collaboration	0 Extérieur	1 Extérieurs	2 Extérieurs	3 Extérieurs	4 Extérieurs	5 Extérieurs	6 Extérieurs	7 Extérieurs	8 Extérieurs	9 Extérieurs et plus	Total
Groupe 1	82	230	324	375	260	209	127	81	56	161	1823
<i>Dont...</i>	<i>(38)</i>										<i>(424***)</i>
...2 CNRS	64	196	268	300	218	172	106	61	47	66	1498
...3 CNRS	15	31	45	60	31	33	15	15	6	7	257
...4 CNRS	3	2	9	12	8	4	6	4	3	4	55
...5 CNRS et plus	0	1	2	3	3	0	0	1	0	3	13
Groupe 2 1 CNRS	*	726	1087	1114	976	708	441	241	128	288	5709 (455**)
Total	82	956	1411	1489	1236	917	568	322	184	367	7532 (493)

* Le nombre d'articles n'ayant qu'un auteur est de 252. Ils correspondent à 132 auteurs qui par ailleurs collaborent.

** Dont 69 chercheurs qui ne publient jamais avec un autre CNRS et qui contribuent pour 697 publications aux articles du groupe 2.

*** Dont 38 chercheurs qui ne publient jamais avec des extérieurs et qui contribuent pour 82 publications aux articles du groupe 1.

Tableau 2 *Intensité moyenne de collaboration inter et intra villes*

	Effectifs	Nombre de villes partenaires (*)	Nombre d'articles en intra	Nombre d'articles en inter (*)	Intensité intra-ville	Intensité inter-ville (moyenne calculée sur l'ensemble des 16 autres villes)	Intensité inter-ville (moyenne calculée sur les partenaires effectifs)
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

Remarque : dans le calcul des intensités, $p=0,0225$.

* Les liens en nombre inférieur à 6 ne sont pas pris en compte - voir section 2.2.

(a) Il s'agit du nombre d'articles pondérés par le nombre de couples participant à leur publication, le nombre d'articles non pondérés étant de 1222.

(b) La somme de cette colonne est de 1530 mais les liens de collaboration sont attribués deux fois (un lien pour chaque partenaire d'un couple), et il faut diviser 1530 par deux pour trouver le nombre total d'articles en inter, soit 765. Il s'agit des articles pondérés par le nombre de couples participant à leur publication, le nombre d'articles non pondérés étant de 412. Au total on retrouve que le champ de l'étude correspond à nombre non pondéré de 1634 articles (=1222+412) – voir section 2.2.

(c) Moyenne simple.

(d) Total en éliminant les doubles comptes. Seulement 34 couples de collaboration se forment sur les 136 couples possibles.

Tableau 3 *Intensité moyenne de collaboration inter et intra laboratoire*

Ville	Nombre de laboratoires par ville	Intensité moyenne de collaboration intra laboratoire	Intensité de collaboration inter laboratoire mais intra ville*	Intensité de collaboration inter laboratoire et inter ville*
Bagneux	1	58,4	-	2,0
Poitiers	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/Moyenne	34	30,7	0,8	0,3

*Moyenne calculée sur l'ensemble des laboratoires

Tableau 4 *Distance géographique moyenne entre les villes partenaires*

	Nombre de laboratoires par ville	Effectifs de chercheurs dans la ville	Nombre de villes partenaires	Distance moyenne aux villes partenaires (en km)
Bagneux	1	9	6	344
Poitiers	1	11	0	0
Gif sur Yvette	1	16	3	171
Grenoble	6	105	12	421
Marseille	1	18	2	361
Meudon	1	9	2	208
Montpellier	3	20	7	548
Orléans	1	10	0	0
Orsay	3	66	9	334
Palaiseau	2	18	4	297
Paris	6	86	7	291
Saint Martin d'Hères	2	31	5	418
Strasbourg	1	14	2	449
Talence	1	9	0	0
Toulouse	2	29	4	410
Villeneuve d'Ascq	1	10	3	305
Villeurbanne	1	9	2	188
Total ou moyenne	34	470	34 ^(a)	279

(a) Total en éliminant les doubles comptes

Tableau 5 *Corrélations entre la distance géographique et l'existence et l'intensité de co-publication à l'échelle des villes et des laboratoires*

Corrélation entre la distance et	Existence de collaboration inter-ville (N=136)	Intensité de collaboration inter-ville (N=34) (a)
	-0.09	-0.16
Corrélation entre la distance et	Existence de collaboration inter-laboratoire (N=561)	Intensité de collaboration inter-laboratoire (N=56) (a)
	-0.18***	-0.002
Corrélation entre la distance et	Existence de collaboration inter-laboratoire <i>inter-ville</i> (N=522)	Intensité de collaboration inter-laboratoire <i>inter-ville</i> (N=41) (a)
	-0.09**	-0.06

*, **, *** significatif respectivement au seuil de confiance de 10%, 5% et 1%.

N représente le nombre d'observations (couples) sur laquelle la corrélation est calculée.

(a) indique que la corrélation est calculée sur les seules observations correspondant à des intensités de collaboration positives.

Annexe 1 : Schéma des deux formes de la collaboration



