



**HAL**  
open science

## Diffusion et structuration spatiale d'une question de recherche en biologie moléculaire

Marion Maisonobe

► **To cite this version:**

Marion Maisonobe. Diffusion et structuration spatiale d'une question de recherche en biologie moléculaire. *M@ppemonde*, 2013, 110 (2), pp.13202. <halshs-01104006>

**HAL Id: halshs-01104006**

**<https://shs.hal.science/halshs-01104006v1>**

Submitted on 15 Jan 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

## Diffusion et structuration spatiale d'une question de recherche en biologie moléculaire

Marion Maisonobe

**Résumé.**— Cet article propose une approche géographique pour étudier l'évolution d'un collectif scientifique de dimension mondiale en biologie moléculaire. Un corpus de 908 publications représentatives d'une question de recherche est analysé. On montre que les résultats obtenus en bibliométrie sont extrêmement sensibles au niveau d'agrégation des données. L'analyse, longitudinale, se centre sur la diffusion spatiale de la question de recherche. Le recours à la théorie des graphes permet d'interroger les effets du temps sur la structure du réseau de copublications et par extension sur l'organisation spatiale du collectif scientifique.

**Agglomérations scientifiques • Bibliométrie • Diffusion spatiale • Réseau de copublications • Théorie des graphes**

**Abstract.**— **Diffusion and spatial structuring of a research question in molecular biology.**— This article is a case study that provides a geographic approach to study the evolution of a scientific community of global dimensions. A corpus of 908 representative publications on a question of molecular biology was analyzed. We show that the results obtained in bibliometrics are extremely sensitive to the aggregation levels of data. The longitudinal analysis focuses on the spatial diffusion of the research question. The use of graph theory helps better examine the effects of time on the structure of the co-publication network and by extension on the spatial organization of the scientific community.

**Bibliometrics • Co-publication network • Graph theory • Scientific centers • Spatial diffusion**

**Résumen.**— **Difusion y estructuración espacial de un problema de investigación en biología molecular.**— Este trabajo presenta un estudio de caso, proponiendo un acercamiento geográfico para el estudio de la evolución de un colectivo científico de tamaño mundial. Se analiza un corpus de 908 publicaciones representativas de un problema de investigación en biología molecular. Se muestra que los resultados obtenidos en bibliometría son muy sensibles al nivel de agregación de los datos. El análisis, longitudinal, se centra en la difusión espacial de la pregunta científica. El recurso a la teoría de grafos permite preguntar los efectos del tiempo sobre la estructura de la red de co-publicación, así como por demás sobre la organización espacial del colectivo científico.

**Aglomeraciones científicas • Bibliometría • Red de co-publicaciones • Teoría de grafos • Difusion espacial**

L'accès en ligne au *Web of Science* permet d'isoler et d'étudier de vastes corpus de publications scientifiques (Cristofoli, 2008) et ainsi d'interroger différentes facettes de l'activité de recherche. Les publications sont un révélateur intéressant bien que partiel du fonctionnement des collectifs scientifiques. En particulier — comme c'est le cas dans cet article — elles sont un moyen de renseigner la dynamique d'un groupe de recherche dans le temps et l'espace.

Si l'on exploite l'information géographique de l'ensemble des publications indexées dans le *Web of Science*, toutes disciplines confondues — par exemple, pour réaliser un classement des villes en fonction de l'importance de leur activité de recherche —

alors les logiques proprement scientifiques s'effacent derrière la hiérarchie des lieux de sciences qui est fonction de leur masse (1). Afin d'échapper à l'effet de taille et d'avoir une meilleure compréhension de l'organisation du travail scientifique, il est intéressant de se rapprocher des contenus, par exemple en se concentrant sur un corpus de publications bien délimité (Vinck, 1999). Cette démarche est assez répandue en *Science studies* — en particulier en sociologie, histoire et anthropologie des sciences. Nous nous proposons de l'appliquer dans une perspective géographique, afin d'analyser, à l'échelle mondiale, l'organisation spatiale de l'activité scientifique dans le cadre d'un problème de recherche particulier.

Nous traitons ici d'une question de recherche en biologie moléculaire qui présente des caractéristiques semblables à l'objet *problem area* décrit par Diana Crane. Pour cette sociologue, « une question de recherche ou *problem area* est un groupement éphémère de chercheurs qui traitent de problèmes spécifiques jusqu'à ce qu'une ou plusieurs décennies plus tard, les problèmes soient résolus ou déclarés insolubles. » (2) (Crane, 1969). À ce niveau, celui où la science se fait, l'organisation sociale, s'il y en a une, est insaisissable, peu structurée et éphémère. Par conséquent, on s'attend à ce que l'organisation spatiale — telle qu'elle est perceptible dans le réseau spatialisé obtenu par l'analyse des copublications — le soit également.

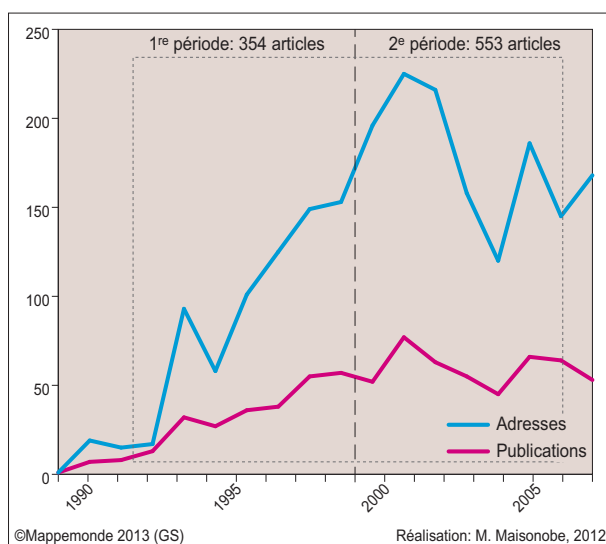
L'objet d'étude est une thématique de recherche qui a regroupé, à partir des années 1990, des scientifiques tâchant de comprendre le lien entre réparation et transcription de l'ADN. Dans un premier temps, la question du niveau d'agrégation géographique des données de base (les contributions scientifiques) est abordée et l'on montre à quel point les résultats y sont sensibles. La diffusion spatiale des publications est ensuite analysée pour rendre compte de l'expansion de la question de recherche dans l'espace mondial. Enfin, l'évolution du graphe de copublications permet de comprendre comment l'expansion du collectif scientifique s'accompagne d'une évolution de sa structure. Il ne s'agit pas seulement d'observer la progression du collectif dans l'espace mais aussi de se demander si cette évolution peut être influencée par des logiques spatiales.

## 1. L'exploitation géographique de données bibliométriques

### 1.1. Description du corpus

Le corpus de publications sur lequel s'appuie cet article a été extrait du *Web of Science*, à partir d'une requête sur le titre, le résumé et les mots clés des notices de la base. Les termes de la requête ont été sélectionnés avec une biologiste moléculaire travaillant sur la question considérée (3). Il s'agit du nom de trois maladies orphelines, dont la maladie de la Lune, et des termes « transcription », « réparation » et « ADN ». Les enfants atteints de ces maladies souffrent d'une déficience dans la réparation des lésions produites sur l'ADN par les rayons UV. En 1993, une équipe spécialisée dans la réparation de l'ADN à Rotterdam et une équipe strasbourgeoise spécialiste de la transcription de l'ADN mettent en évidence l'existence d'un lien entre la réparation des lésions dues aux UV (ou réparation par excision de nucléotide) et la transcription. C'est à partir de cette découverte que se multiplient les travaux pour caractériser le lien entre réparation et transcription de l'ADN en se concentrant sur l'une ou plusieurs de ces trois maladies.

Les publications traitant de cette question paraissent à partir de 1991 et leur nombre décolle avec la découverte de 1993. Pour une analyse longitudinale, le corpus a été découpé en deux périodes de même durée (7 à 8 ans): 1991-1999 et 2000-2007. La première période est une phase d'émergence et d'émulation scientifique au cours de laquelle sortent 40 % des publications. Le nombre d'articles progresse en début de deuxième période jusqu'à atteindre son maximum en 2001, puis il baisse et stagne au niveau de 50 articles par an, ce qui laisse supposer que la question de recherche s'épuise (fig. 1).



**1. Nombre d'adresses par publication et de publications par an**

Le passage d'un stade de recherche à un autre est une clé indispensable pour interpréter l'évolution d'un groupe scientifique. Les résultats de la célèbre étude de Nicholas Mullins (1972) sur le groupe Phage (groupe pionnier en biologie moléculaire qui, à partir de 1935, a été à l'origine de la structuration d'un vaste champ de recherche) allaient déjà dans ce sens. Il distinguait quatre stades d'évolution depuis l'émergence jusqu'à la stabilisation du groupe sous forme de spécialité scientifique (4). À la différence du champ étudié par Nicholas Mullins, notre *problem area* ne correspond qu'à une seule question de recherche. Le sujet ici est investi en majorité par des membres de la « communauté de la réparation de l'ADN » (et plus spécifiquement par le groupe Nucleotide excision repair) et dans une moindre mesure par des membres de la « communauté de la transcription de l'ADN » (comme l'équipe strasbourgeoise). Le collectif étudié est hybride. Il n'est pas institutionnalisé et on imagine qu'il peut se dissoudre rapidement. Ce type d'évolution est connu mais la traduction spatiale d'un tel processus l'est beaucoup moins.

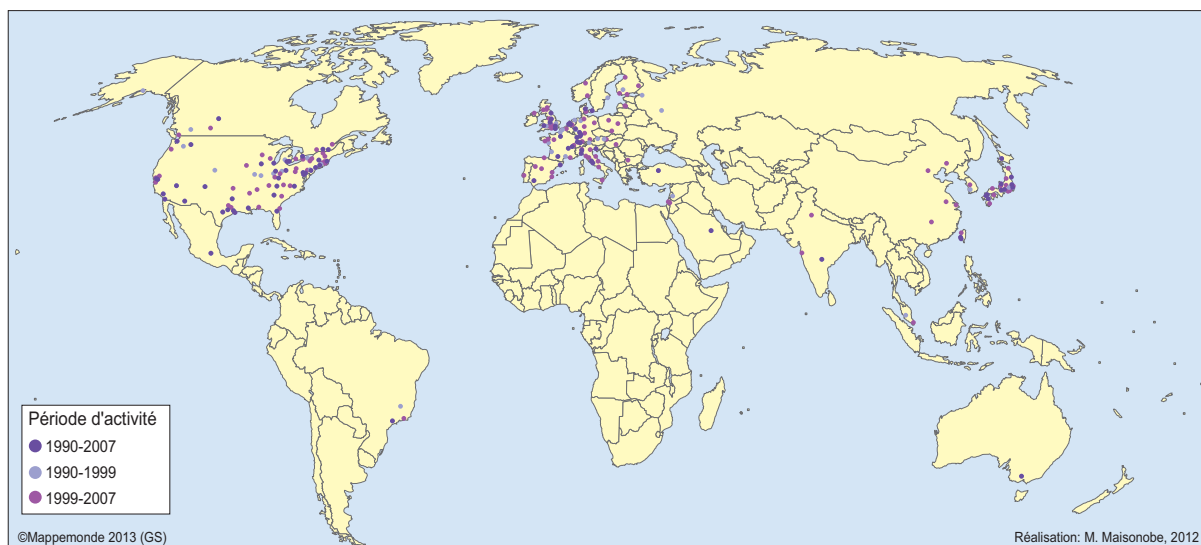
## 1.2. Méthodologie

À chaque publication extraite de la base sont associées les adresses ou affiliations des auteurs. L'exploitation du champ « adresse » permet de localiser les activités de publications (fig. 2). En complément, les cosignatures d'articles permettent de saisir la structure du groupe et d'interroger sa cohésion. Dans un réseau de coauteurs, on considère classiquement que les nœuds sont les auteurs et que deux auteurs sont connectés par un lien s'ils ont coécrit un ou plusieurs articles (Newman, 2004).

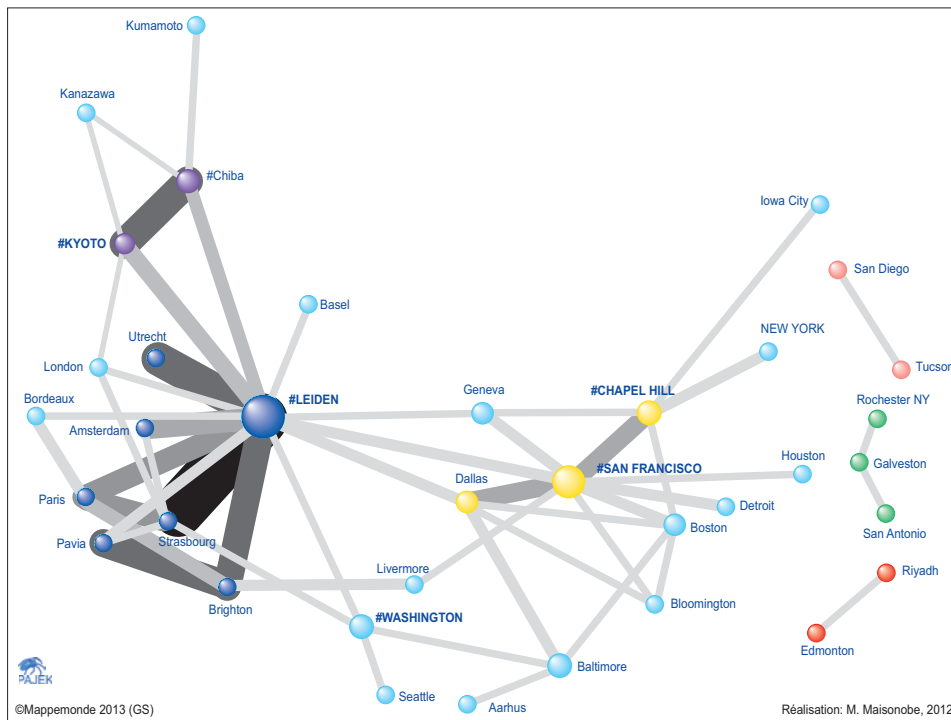
Ici, le réseau est abordé sous un autre angle : les publications des auteurs appartenant à une même entité géographique sont agrégées (fig. 3a et 3b). Ainsi, les nœuds sont des entités géographiques et les liens entre deux entités sont valués selon le nombre de publications cosignées par ces entités. Cette approche permet de situer directement l'analyse dans le registre de la géographie : on identifie les articles de l'échantillon par le lieu de production et non par les auteurs, ce qui fait qu'on analyse un réseau de lieux et non d'individus.

Le recours à la théorie des graphes permet de situer et de caractériser les lieux de publications les uns par rapport aux autres en fonction de l'intensité de leurs relations. Tandis que sur la figure 2, c'est la distance physique (kilométrique) qui permet de les situer ; sur les figures 3a et 3b, c'est l'intensité relationnelle qui sert de métrique (5). Les deux approches sont complémentaires pour appréhender les lieux en situation (leur position relative et leur évolution).

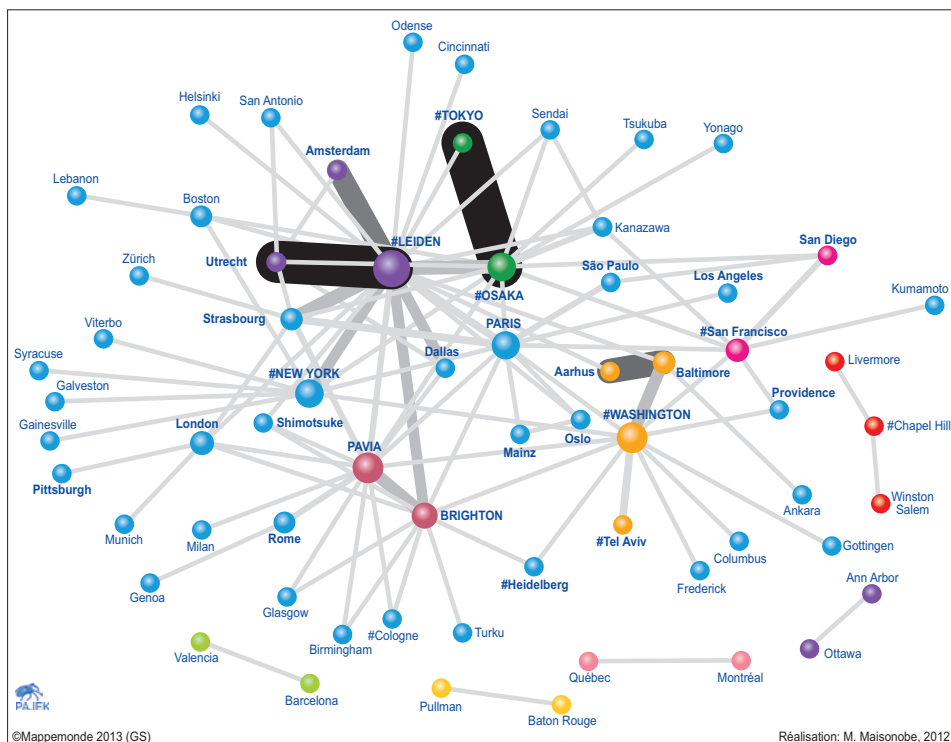
Pour cette étude, plusieurs niveaux géographiques d'agrégation des données ont été considérés (Baron *et al.*, 2011). Avec la méthode de délimitation des agglomérations urbaines retenue ici (Eckert *et al.*, 2013), le corpus étudié passe de 298 villes à 229 agglomérations. Ce regroupement en agglomérations peut poser quelques problèmes, par exemple pour la baie de San Francisco, où tel ou tel choix d'agrégation spatiale modifie considérablement l'allure du réseau de copublications.



## 2. La répartition des agglomérations publiantes selon leur période d'entrée dans la question de recherche



**3a. 1991-1999: Leiden-Rotterdam au centre**



**3b. 2000-2007: Une organisation polycentrique**

**3. L'évolution de la structure du réseau de copublications dans le temps**

Graphes des liens supérieurs ou égaux à deux. Les groupes de couleurs (sauf turquoise) correspondent à des sous-groupes connexes cohésifs ou « Islands ». La surface des nœuds dépend de l'importance de leur indice d'intermédierité. Les agglomérations dont le label est précédé du signe « # » sont le produit du regroupement de plusieurs villes publiantes en agglomération.

### 2.3. L'influence des découpages géographiques sur les résultats

Les premiers articles cosignés par des chercheurs issus de plusieurs agglomérations sont publiés en 1991 et les premières collaborations internationales apparaissent en 1993. Comme on peut le voir sur la [figure 1](#), le nombre d'adresses par article augmente jusque dans les années 2000 (on passe en moyenne de 2 à 2,5 adresses par article) ce qui se traduit par une augmentation limitée du nombre moyen d'agglomérations distinctes étant à l'origine d'un article cosigné (de 1,65 à 1,76). Le regroupement des données au niveau des agglomérations a un effet direct sur la taille du réseau de copublications. Un grand nombre de cosignatures est ignoré puisque les collaborations internes à une institution et à une agglomération ne sont pas prises en compte (6). Il reste que bien souvent (pour 70 % des agglomérations ici) parler des publications d'une agglomération revient à parler des publications de l'unique groupe de recherche intéressé par cette question dans cette agglomération. Seules 70 agglomérations (soit 30 %) regroupent plusieurs affiliations différentes. Les publications du corpus sont d'ailleurs extrêmement concentrées. Ici la « loi de Lotka » s'applique : 80 % des publications sont produites par 5 % des agglomérations, et donc les zones à forte densité de publications sont rares ([encadré 1](#)). Dès lors, il n'y a qu'un petit nombre de zones particulièrement sensibles aux choix faits en matière de regroupement en agglomérations. Les cas délicats concernent les lieux caractérisés à la fois par une forte densité de publications et d'affiliations. Le rang (en nombre de publications) des quelques lieux situés en zone dense, mais surtout leur position dans le réseau de copublications, est sensible aux découpages géographiques retenus.

#### Encadré 1. La loi Lotka

Tout comme les lois de Pareto et Zipf (1935), la loi de Lotka (1926) est une loi de puissance. Elle établit que le nombre d'auteurs publiant  $i$  articles est inversement proportionnel au carré d'articles publiés. Donc, ce nombre est égal à :

$$n_i = n_1 / i^2, \quad i = 1, 2, 3, \dots, i_{\max}$$

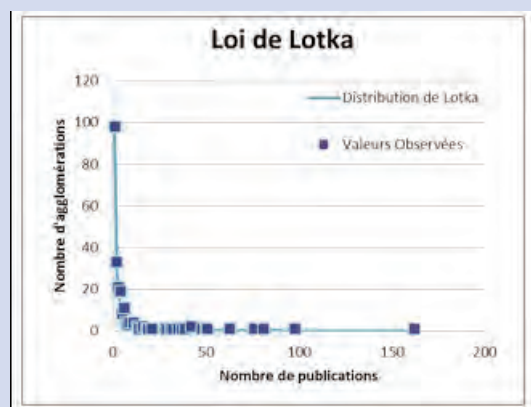
Où  $n_1$  est le nombre de scientifiques ayant écrit un article et  $i_{\max}$  le nombre maximum d'articles écrits par un auteur (Polanco, 1995 ; Solla-Price, 1963). Cette loi traduit le fait que l'essentiel de la production est le fait d'une petite poignée de publiants.

Les scientomètres ont largement recours à cette loi pour expliquer ou comparer les distributions très asymétriques auxquelles ils ont à faire. Cependant, l'exposant 2 retenu par Lotka ne se retrouve pas souvent dans la réalité. Il implique qu'environ 60% des auteurs d'un échantillon donné ne publient qu'un seul article. Les pratiques de publications ayant évolué :  $i_{\max}$  tend à être plus grand et  $n_1$  plus petit. Pour notre cas d'étude, on montre que la relation existe, y compris si l'on prend comme variable à expliquer la productivité des agglomérations et non celle des auteurs. Ainsi, 98 agglomérations (40%) n'ont publié qu'un seul article. Par conséquent,

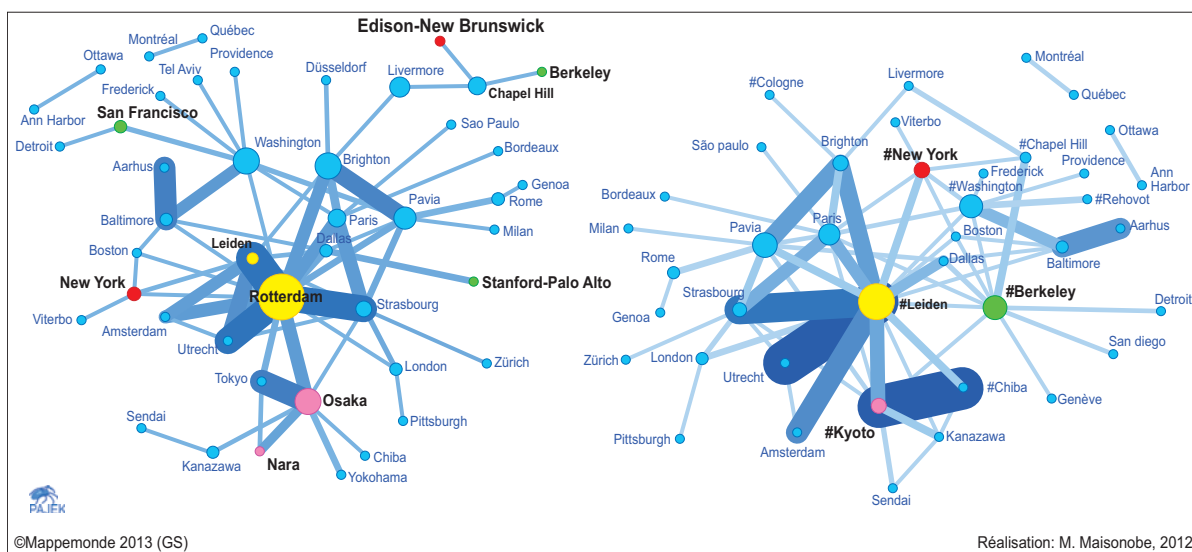
l'exposant le mieux adapté pour décrire la relation est de 1,5 ; ce que confirment par ailleurs les résultats du test d'adéquation de Kolmogorov-Smirnov. On obtient la fonction suivante :

$$Y = b / X^a \quad \text{où } b=98 \text{ et } a=1,5$$

Ainsi, 5 % des agglomérations sont responsables de 80 % de la production et ce ratio est proportionnel aux agglomérations n'ayant publié qu'un seul article.



Pour en témoigner, la **figure 4** montre deux représentations du graphe simplifié des copublications. Sur la représentation de gauche, les entités sont les villes et leur proche banlieue (premier niveau d'agrégation: 45 nœuds). À droite, le même graphe a été représenté à partir des agglomérations obtenues grâce à la méthode finalement retenue (second niveau d'agrégation: 40 nœuds). La surface des nœuds est proportionnelle à l'*Indice de centralité d'intermédiarité* (7) qui permet de mettre en évidence la qualité d'entremetteur occupée par les nœuds dans le réseau. Deux choix d'agrégation ont un effet considérable sur la structure du réseau: le regroupement de Stanford, San Francisco et Berkeley (en vert) et de New York avec Edison et Brookhaven (en rouge). Ces regroupements améliorent artificiellement la connexion des agglomérations américaines qui étaient auparavant dispersées. Avec ce nouveau découpage, l'agglomération de New York (#New York) et la baie de San Francisco (#Berkeley) prennent une place centrale dans le réseau. La pertinence de ces deux regroupements peut poser problème pour notre analyse car les localités publiantes qui y sont regroupées n'entretiennent pas de collaborations privilégiées en biologie moléculaire. En revanche, le fait que Leiden et Rotterdam aux Pays-Bas (en jaune) d'une part, et qu'Osaka et Nara au Japon (en violet) d'autre part, ne forment plus qu'une entité n'a pas tellement d'influence sur la structure globale du graphe. Compte tenu de l'intensité de leurs collaborations, leur regroupement est à la fois cohérent géographiquement et scientifiquement.



#### 4. L'effet sur la structure du réseau de copublications (1990-2007) de l'agrégation des données à différents niveaux géographiques.

Les liens avec une valeur inférieure à trois ont été supprimés et les isolats ne sont pas représentés. à gauche: les villes; à droite: les agglomérations précédées du signe #

Les collaborations scientifiques étant un bon indicateur de la structuration de la question de recherche au niveau international, le fait que deux entités géographiquement proches ne copublient pas est une information importante sur la cohésion scientifique locale. Ainsi, au lieu de s'en tenir à la cohérence géographique, une autre possibilité en matière de découpage des agglomérations serait de prendre en considération la cohérence scientifique des regroupements effectués, mais cela pose plusieurs problèmes. Les données relationnelles du corpus étudié ici ne peuvent être mobilisées pour légitimer le découpage des agglomérations. C'est, rappelons-le,

un petit corpus et on y repère peu de relations de complicité scientifique fortes (d'entités ayant cosigné ensemble au moins deux articles sur la question). En effet, seulement 2 liens sur 10 dans le réseau de copublications ont une valeur supérieure à 1. De plus, pousser ce raisonnement jusqu'au bout conduirait à remettre en cause certains regroupements intra-institutionnels ou intra-municipaux qui ne seraient pas forcément sanctionnés par une collaboration scientifique. Puisque c'est le processus de diffusion spatiale de la question de recherche qui est placé au centre de l'analyse, alors il est primordial que les entités considérées soient cohérentes géographiquement (**encadré 2**). On pourrait imaginer une méthode permettant de coupler cohérence géographique et scientifique mais il faudrait pour cela mobiliser des données plus générales et significatives sur les collaborations que celles relatives à la communauté de recherche que nous étudions ici.

### **Encadré 2. Cohérence géographique des regroupements versus cohérence scientifique, un exemple américain tiré du corpus**

Le National Cancer Institute (NCI) travaille beaucoup avec le National Institute of Aging (NIA) sur le lien entre transcription et réparation de l'ADN. Ces deux institutions partagent une proximité organisationnelle puisqu'elles sont chacune rattachées au National Institute of Health (NIH). Au début des années 1990, elles sont toutes deux situées à Bethesda (dans la banlieue de Washington) si bien que, quel que soit leur degré de collaboration, elles appartiennent à la même entité géographique (l'agglomération de Washington). Au milieu des années 1990, le NIA déménage à Baltimore ce qui se traduit par la mise en place de collaborations régulières entre Baltimore et Bethesda. Un système de cours par vidéoconférence est lancé à cette occasion qui ne tardera pas à être rendu accessible aux autres centres NIH du territoire américain puis mis en ligne pour un accès généralisé à tous (Kraemer, Bohr, 2005). Ce déménagement

suffit-il à justifier le regroupement de Baltimore et Washington dans une même entité ? Si le NIA avait déménagé à l'autre bout des États-Unis, l'intensité de la collaboration n'aurait pas forcément été différente puisque les deux centres bénéficient d'une forte complicité scientifique et de moyens technologiques facilitant leurs échanges.

Si la problématique était celle de la structuration scientifique de la question de recherche (sans considération pour son organisation spatiale), il aurait pu être pertinent de regrouper ensemble tous les centres du NIH dans une même entité (quelle que soit leur localisation). Cela aurait permis de situer le NIH par rapport aux autres organisations scientifiques travaillant sur ce sujet de recherche ; mais n'aurait pas eu de sens dans le cadre d'une étude comme celle-ci qui prend la localisation géographique des activités scientifiques comme variable pour analyser leur organisation.

## **3. Le groupe « DNA Transcription and Repair » : son expansion et sa cohésion**

### **3.1. La diffusion spatiale de la question de recherche**

Sur la **figure 2** sont représentées les agglomérations actives dans l'échantillon : en gris-bleu, celles qui ne sont actives qu'au cours de la première période (32 agglomérations). En violet, apparaissent les 87 agglomérations qui publient également dans la seconde période et en fuchsia, celles qui ne publient que pendant la seconde période (110).

Les plus fortes densités d'agglomérations publiantes se trouvent autour des centres pionniers de la question de recherche : aux États-Unis (sur la côte Est américaine, en Californie et au Texas), au cœur de l'Europe autour du triangle qui a été à l'origine même de la question de recherche (Leiden-Rotterdam-Strasbourg) et au Japon. À l'exception de la Chine qui n'apparaît qu'à partir de la seconde période, la diffusion spatiale se fait principalement par « contagion » au sein des pays les plus productifs de la première période (8) (Saint-Julien, 2004). La question de recherche est suffisamment porteuse, et les résultats scientifiques obtenus au cours de la

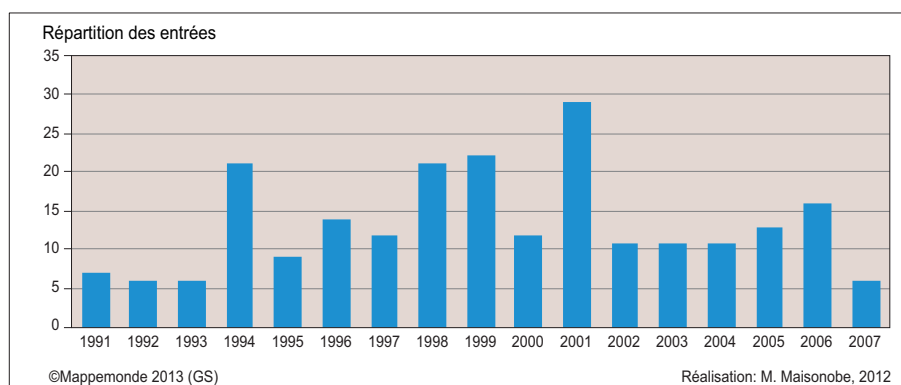
première période assez révélateurs, pour justifier un « rattrapage » de certains pays au cours de la seconde période (Italie, Espagne et Finlande comptent alors davantage d'agglomérations impliquées). La question de recherche s'étend ainsi à onze nouveaux pays dont cinq en Europe de l'Est (9). Le nombre total de pays concernés reste cependant relativement limité (36) en raison notamment du degré de spécialisation requis dans le domaine de la réparation de l'ADN (10).

On l'a vu, la majeure partie des publications de l'échantillon est produite par un nombre restreint d'agglomérations. En raison de la nature du *problem area*, à côté d'un noyau dur d'équipes réellement engagées dans la question de recherche (qui publient régulièrement) se trouvent quantité d'équipes dont la participation aux travaux reste épisodique (un ou deux articles). Sachant que peu d'équipes investissent cette question, on peut distinguer plusieurs catégories d'adoptants à la manière de Everett Rogers (1962), « les innovateurs et primo-adoptants » et le « adoptants tardifs » (la majorité).

Une manière de formaliser cette distinction a été de noter systématiquement l'année du premier article publié, pour chaque agglomération, puis le nombre de fois, en nombre d'années, où l'agglomération produit à nouveau sur le sujet. Quand ce dernier nombre est rapporté au nombre potentiel d'années, c'est-à-dire au nombre d'années restantes (dans la période considérée) à partir de la sortie du premier article, on obtient un indice d'« investissement » de l'agglomération. Cet indice, extrêmement sensible au temps, est délicat à utiliser : plus l'entrée est tardive dans la période, moins l'indice est significatif. Dans la première période, 32 agglomérations ont un indice supérieur à 0,5, c'est-à-dire qu'elles sont présentes plus de la moitié du temps espéré compte tenu de leur année d'entrée. Il n'y en a que trois dans ce cas-là pendant la seconde période (Oslo, Varsovie et Pittsburgh). Les années les plus dynamiques en termes de production sont celles qui se trouvent à la charnière entre les deux périodes étudiées, ce qui explique en partie les résultats obtenus : 74 % des agglomérations entrées avant 1999 republient au moins une fois avant 1999 (elles sont 80 % à republier au moins une fois avant 2007) quand seulement 27 % des agglomérations entrées après 2000 republient au moins une fois avant 2007.

Cette évolution s'explique par l'épuisement de la question de recherche. Les nombreux entrants durant la seconde période (fig. 2 et 5) sont des « adoptants tardifs » qui, en majorité, ne s'investissent pas véritablement dans le sujet. Les cartes d'évolution de stocks de publications permettent d'en rendre compte (fig. 6). On observe encore que les agglomérations entrées tardivement apparaissent à proximité des agglomérations pionnières. C'est très net pour les agglomérations les plus « innovantes » qui ont fait paraître au moins cinq publications au cours de la seconde période (leur nom est mentionné sur les cartes détaillées). On repère sur la carte américaine l'exception que constitue Cuernavaca (Mexique), et, sur la carte européenne, l'exception norvégienne d'Oslo.

Non seulement il n'est pas facile de s'approprier les méthodes et le matériel (y compris les cellules) nécessaires à l'investigation mais en plus, la question a eu tendance à s'épuiser à partir du moment où tous les gènes impliqués dans le lien entre réparation et transcription de l'ADN ont été clonés. Ce sont les éléments d'explication suggérés par les données, et confirmés par plusieurs biologistes travaillant sur cette question, qui permettent d'interpréter la timidité de ce processus de diffusion.



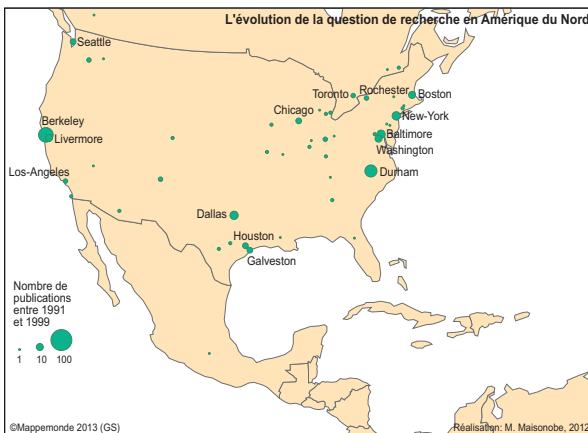
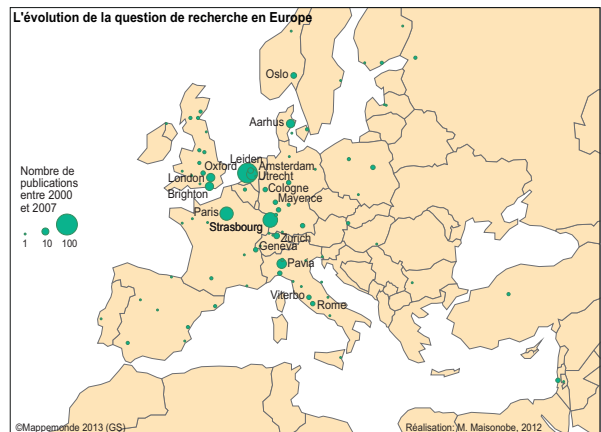
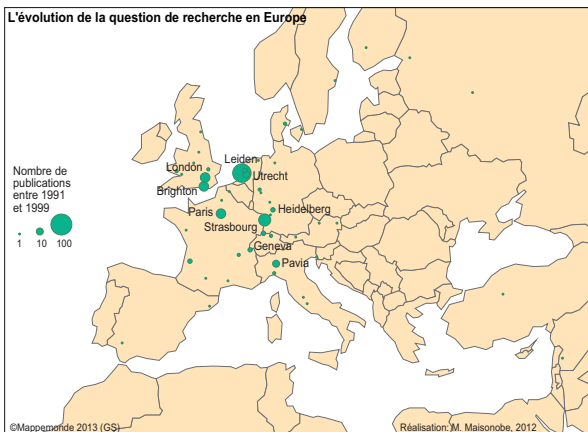
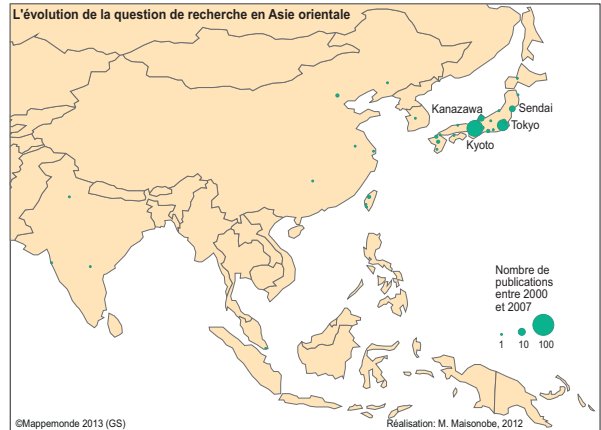
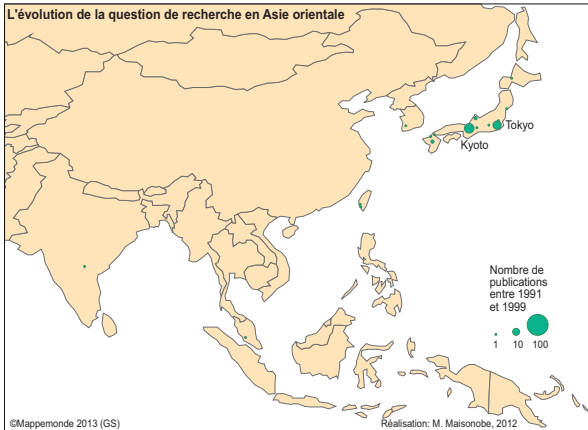
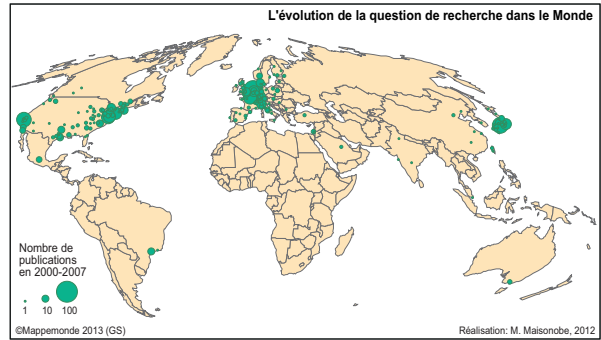
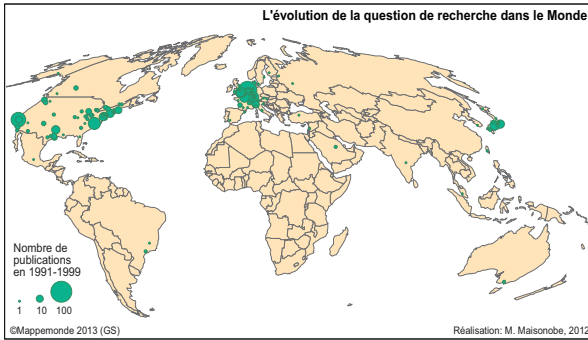
### 5. Nombre annuel de nouvelles agglomérations publiantes sur la période étudiée

### 3.2. Une « autonomie scientifique » inégale vis-à-vis de la question de recherche

Dans notre étude, 45 % des publications ont été cosignées par au moins deux agglomérations différentes : 147 dans la première période (41 %) et 263 dans la seconde (47 %). Alors que la procédure d'extraction des données ne le laissait pas prévoir, le réseau de copublications obtenu à partir du corpus compte très peu de composants non connectés à la composante principale : un seul dans la première période et quatre dans la seconde période qui correspondent à des collaborations très locales. Aussi, sur les 230 agglomérations considérées, il n'y en a que 25 qui soient isolées (11).

Pourquoi ? Pour investir cette question de recherche, il faut pouvoir se mettre en relation avec des centres disposant de cellules spécifiques, celles de patients présentant les maladies orphelines étudiées. En Europe, les bibliothèques de cellules se trouvent à Rotterdam, Paris, Pavie et Brighton. Le sociologue Dominique Vinck l'a montré en étudiant un programme européen de santé : les réseaux scientifiques ne peuvent se comprendre qu'en tenant compte de l'importance des objets intermédiaires — comme ici, les bibliothèques de cellules — dans la construction et l'organisation des collectifs (Vinck, 1999). L'importance cruciale de l'accès aux cellules et à certains savoirs-faire explique les mises en relation des différents centres travaillant cette question dans le monde, d'où le petit nombre d'isolats repéré. La centralité de certaines équipes (et donc, pour nous, de certaines villes) est liée au fait qu'elles disposent des bibliothèques de cellules dont toutes les autres ont besoin.

Les isolats ayant un niveau de production très bas (un ou deux articles publiés pour toute la période) ont un poids très faible dans le corpus. Ce sont des cas particuliers puisque dans le cadre de cette question de recherche, moins l'on est productif sur le sujet, plus il y a de chance que nos productions soient des coproductions : 80 % des 98 agglomérations qui n'ont publié qu'un seul article pour toute la période, l'ont publié en collaboration avec une autre agglomération. C'est un effet de la moindre autonomie de ces agglomérations. Ce sont les agglomérations les plus productives qui ont aussi tendance à produire le plus de publications en interne (non cosignées avec une autre agglomération). Être autonome sur la question suppose d'avoir le matériel et le savoir-faire à disposition : 70 % des 500 publications publiées en interne ont été produites par les 28 agglomérations les plus productives (ayant signé plus de 10 publications au total) (12). Il y a quelques contre-exemples : Pavie, Aarhus, Amsterdam et Kanazawa, parmi les agglomérations les plus productives, ne publient jamais seules.



## 6. L'évolution de la question de la recherche dans le monde

Les articles recensés à Aarhus (Danemark) sont en fait produits sous le patronage d'A. Bohr, chercheur qui a une double affiliation, danoise et étatsunienne : l'université d'Aarhus et le National Institute of Aging de Baltimore. Les autres cas sont moins évidents à expliquer mais concernent des agglomérations qui sont intégrées dans des structures de collaboration denses : Pavie entretient avec Brighton une relation privilégiée sur ce sujet (les équipes qui y sont situées ont des programmes de recherche communs), Amsterdam est proche des initiateurs néerlandais de la question de recherche et n'est donc pas très autonome vis-à-vis d'eux (Rotterdam-Leiden) et, de la même manière, Kanazawa s'inscrit dans le cluster japonais.

Autre point notable : les collaborations intra-nationales sont généralement le fait des pays les plus productifs. Au cours de la première période, il n'y a que huit pays dont les agglomérations collaborent au niveau national et ce sont les huit premiers producteurs. On trouve cinq nouveaux pays autonomes dans la seconde période (13). Quatre pays sont tout aussi actifs pour la même période mais ne disposent pas d'un réseau d'agglomérations autonomes : le Danemark dont le bon niveau de publication s'explique par les relations avec Baltimore ; la Norvège et la Pologne avec une activité forte d'agglomérations « solitaires » entrées tardivement sur le sujet (Oslo et Varsovie déjà évoquées) ; et enfin, le Mexique. Les productions mexicaines sont toutes réalisées à Cuernavaca : l'activité y démarre en 1999 et reste régulière et autonome jusqu'en 2007 où l'on note, enfin, une collaboration avec l'équipe de la transcription de Strasbourg.

### 3.3. Le passage d'une logique d'innovation à une logique de diffusion

L'évolution des liens de copublication témoigne, comme la logique de diffusion spatiale, du passage d'une période d'émulation scientifique — d'innovation — à une période de diffusion — d'exploitation — des résultats. On constate qu'à l'exception des agglomérations étatsuniennes, celles du collectif collaborent moins avec l'étranger dans la seconde période qu'elles ne le faisaient au cours de la première.

Si l'on considère l'ensemble des liens, il semble que la part des liens intra-nationaux diminue dans le temps en nombre et en valeur. En effet, d'une période à une autre, le réseau se densifie et s'agrandit. Pris dans sa globalité, il paraît s'internationaliser en raison de l'ouverture des États-Unis au cours de la seconde période. En valeur et en nombre, la part des liens « intra-nationaux » sur l'ensemble des liens du réseau (hors États-Unis) progresse d'une période à une autre (14). En fait, les États-Unis sont omniprésents dans le réseau de copublications ce qui influe largement sur l'allure d'ensemble du réseau (les liens entre deux entités dont l'une au moins est étatsunienne représentent en nombre 64 % des liens de la seconde période et 56 % en valeur). Or, d'une période à une autre, la part des collaborations intra-nationales augmente en valeur pour tous les pays autonomes de la première période à l'exception des États-Unis (15). Dès le départ, les différents groupes américains (Washington-Baltimore, la côte californienne, le groupe texan, l'agglomération new-yorkaise) sont peu connectés, ce qui suggère l'existence d'une compétition à l'intérieur des frontières étatsuniennes pour le partage des données et des résultats. Dans la seconde période, on observe une ouverture des agglomérations étatsuniennes à l'international qui ne s'accompagne cependant pas d'une intensification de leurs relations intra-nationales. Dans les autres pays, l'augmentation des liens intra-nationaux est sensible pour le Japon, les Pays-Bas, le Royaume-Uni et l'Italie. Cette évolution ne s'explique pas seulement par l'entrée de

nouvelles équipes sur le sujet à proximité des pionnières (suivant le principe de contagion), mais aussi par la diminution en valeur de liens internationaux qui furent, pendant un temps, structurants pour la question de recherche.

Dans ce réseau, comme on pouvait s'y attendre, les liens sont particulièrement instables ou éphémères. Alors que 87 agglomérations sur les 119 pionnières republient après l'an 2000, seulement 77 des 271 liens de copublications enregistrés pendant la première période se maintiennent dans la seconde. Ces liens structurent le réseau de copublications : 83 % de l'évolution de la valeur de ces liens s'explique par la progression de la valeur des liens « intra-nationaux ».

Le graphe de la première période montre une structure dont la hiérarchie est lisible. Il y a un groupe très cohésif autour des instigateurs de la question de recherche (le groupe européen) auquel se connecte d'un côté, le groupe japonais par l'intermédiaire d'Osaka et de l'autre, les États-Unis, qui constituent un groupe moins cohésif (les trois relais principaux sont Washington, la Baie de San Francisco et Research Triangle Park (#Chapel-Hill en Caroline du Nord). Le graphe de la seconde période est plus complexe. Il y a davantage de liens entre les membres des différents groupes : la structure d'ensemble est moins hiérarchisée. Paris, New York et Pavia deviennent aussi centrales que l'étaient Leiden et Washington dans la première période (avec un indice de centralité de proximité supérieur à 0,5, 16). Le groupe néerlandais tend à se refermer sur lui-même : il forme une « île » à part, c'est-à-dire, un sous-groupe connexe dont les nœuds sont davantage connectés entre eux qu'avec leurs voisins (17) (Batagelj *et al.*, 2006). Au fil du temps, les équipes de Rotterdam et Leiden tendent à moins travailler avec celle de Strasbourg ce qui s'explique par l'épuisement de la question de recherche à l'origine de laquelle on trouve ce trio. Ce lien se justifiait parce que la thématique requérait la compétence de spécialistes de la transcription de l'ADN. Une fois le problème résolu, l'équipe de Rotterdam a eu moins besoin de ce lien, crucial à l'origine. Cette interprétation a été validée par les spécialistes néerlandais du domaine.

Ainsi, il apparaît que les équipes leaders des pays autonomes contribuent à la diffusion de la question de recherche par contagion tandis qu'elles interagissent de moins en moins avec l'étranger. L'intérêt stratégique de certaines collaborations tend à diminuer à mesure que s'épuise le potentiel d'innovation de la question de recherche. L'étude de cas d'Atkinson *et al.* (1998) sur le clonage d'un gène réalisé par plusieurs groupes de recherche en 1991 montre bien comment les relations mises en place au départ (au moment le plus stratégique) tendent à s'émousser lorsqu'elles ne sont plus nécessaires. L'idée d'Atkinson est qu'au partage initial succède la compétition, le « chacun pour soi ». Cette étude portant sur une équipe spécifique est éclairante pour le cas qui nous intéresse.

L'analyse des graphes permet d'observer que la dynamique de recherche fondée pendant la première période sur l'entretien de relations internationales entre équipes innovantes (18) s'essouffle dans la seconde période au profit de logiques relationnelles relevant d'autres types de mécanismes comme l'« homophilie » (19) ou la « triadic closure » (le fait de développer des relations avec les proches de ses proches) (Powell *et al.*, 2005 ; Ter Wal, Boschma, 2008). On rejoint les résultats de Leydesdorff et Rafols qui, comparant l'évolution des réseaux de copublications pour deux technologies émergentes, montrent que, pour les deux technologies, la logique d'attachement préférentiel tend à diminuer au profit d'une logique de petit-monde

(Leydesdorff, Raffols, 2011). Ce changement de stratégie est compréhensible si l'on tient compte du passage d'un stade de recherche à un autre. L'enjeu scientifique n'est plus le même dans la seconde période puisqu'à partir du moment où tous les gènes ont été clonés, il est davantage question de diffusion que d'innovation.

## Conclusion

On a pu montrer que le passage d'un stade de recherche à un autre dans le cadre de l'évolution d'une question scientifique a une influence sur sa diffusion dans l'espace géographique ainsi que sur l'organisation spatiale de son réseau de collaboration. La diffusion de la question s'organise essentiellement au voisinage géographique des pionniers. S'il y a de nouveaux pays qui participent au cours de la seconde période, leur niveau de participation et son intensité ne témoignent pas, en général, d'un engagement fort mais plutôt d'une prise de connaissance de la question scientifique et éventuellement d'une exploitation des résultats. Ainsi, du fait de l'épuisement progressif de cette question, la diffusion reste relativement limitée dans l'espace géographique, d'autant plus que le sujet n'est pas aisément appropriable pour un nouveau venu. Il y a un coût à l'entrée qui ne se limite pas à acquérir le savoir-faire mais suppose aussi de se procurer le matériel et les données nécessaires. Ce dernier requisit exige d'être en contact avec une équipe disposant d'une bibliothèque de cellules. Pour cette raison, les productions « autonomes » sont essentiellement les principaux contributeurs qui ont à leur disposition tout ce qui est nécessaire pour l'avancement des recherches. Ainsi, le réseau va en se densifiant et rares sont les agglomérations qui n'y sont pas connectées à un moment donné.

Si des travaux antérieurs ont permis d'établir que la structure d'un réseau scientifique est sensible au passage d'un stade de recherche à un autre (Mullins, 1972; Atkinson *et al.*, 1998; Leydesdorff, Raffols, 2011), ce cas permet d'explorer plus précisément comment évolue la structure spatiale d'un tel réseau. Il va dans le sens des travaux précédents qui montrent qu'une fois la question moins brûlante, les collaborations stratégiques ou opportunistes se révèlent moins indispensables et tendent à diminuer au profit de relations de complicité. Ainsi, à mesure que la question s'épuise, en Europe comme au Japon, les agglomérations se replient sur leur contexte national et l'intensité de leurs relations internationales décroît considérablement. À rebours de tous les autres pays, les États-Unis ont tendance à s'ouvrir ce qui donne l'apparence, de prime abord, qu'un mouvement général d'internationalisation affecte le réseau. Ce résultat atteste du fait que le niveau national est toujours pertinent pour comprendre la structuration des collectifs scientifiques et ne devrait pas être éliminé des réflexions portant sur les réseaux scientifiques (Arvanitis, 2011).

Comme pour l'étude de Loet Leydesdorff et Ismael Raffols, l'analyse se fonde sur l'exploitation de données bibliométriques et en particulier sur le traitement de l'information géographique qui y est contenue. Cependant, en se limitant à une question de recherche bien définie et circonscrite, il a été possible de tenir compte des objets intermédiaires pour interpréter les évolutions observées. En outre, l'agrégation des données à un niveau permettant une comparaison internationale des lieux de publication constitue un apport notable aux travaux antérieurs qui ne prenaient pas ce problème en considération. Le recours à la statistique descriptive permet d'observer des cas particuliers. On pourrait prolonger ce travail par une

analyse de réseaux multinationaux afin de repérer des régularités ou profils spécifiques (grande agglomération dans un grand pays *versus* petit pays, petite agglomération dans un grand pays *versus* petit pays) selon la méthode du *linked-design structural* (Lazega *et al.*, 2007).

Cette étude montre qu'il y a un réel intérêt à considérer l'espace comme variable pour analyser l'évolution des collectifs scientifiques. Elle conforte l'idée que les groupes scientifiques ne sont pas stables dans le temps et gagnent à être étudiés en dynamique. Elle appelle à être vigilant quant à la sensibilité des données aux regroupements selon des critères géographiques.

## Bibliographie

- ARVANITIS R. (2011). « Que des réseaux! Compte rendu de Caroline Wagner. The New Invisible College. Science for Development ». *Revue d'anthropologie des connaissances*, vol. 5, n° 1, p. 178-185. doi:10.3917/rac.012.0178
- ATKINSON P., BATCHELOR C., PARSONS E. (1998). « Trajectories of collaboration and competition in a medical discovery ». *Science Technology and Human Values*, vol. 23, n° 3, p. 259-284. doi:10.1177/016224399802300301
- BARON M., ECKERT D., MAISONOBE M. (2011). « Worlds of Scientific Cooperation. Choosing appropriate data for specific networks ». *17th European Colloquium on Quantitative and Theoretical Geography*, Athènes, 2-5 septembre 2011.
- BATAGELJ V., KEJZAR N., KORENJAK-CERNE J. (2006). « Analyzing the structure of U.S. patents network ». In BATAGELJ V., BOCK H.-H., FERLIGOJ A., ZIBERNA A. *Data Science and Classification*. Heidelberg: Springer, 358 p. ISBN: 3-540-34415-2
- BEAUGUITTE L. (2011). *L'Assemblée générale de l'ONU de 1985 à nos jours: acteur et reflet du Système-monde. Essai de géographie politique quantitative*. Paris: Université Paris Diderot, thèse de doctorat de géographie, 311 p. <http://tel.archives-ouvertes.fr/tel-00634403> ; oai:tel.archives-ouvertes.fr:tel-00634403
- CRANE D. (1969). « Social structure in a group of scientists: A test of the "Invisible College" Hypothesis ». *American Sociological Review*, vol. 34, n° 3, p. 335–352. <http://www.jstor.org/stable/2092499>
- CRISTOFOLI P. (2008). « Aux sources des grands réseaux d'interactions ». *Réseaux*, vol. 6, n° 152, p. 21-58. doi:10.3917/res.152.0059
- DUCRUET C., IETRI D., ROZENBLAT C. (2011). « Cities in worldwide air and sea flows: a multiple networks analysis ». *Cybergeo*, article n° 528, p. 1-23. <http://cybergeo.revues.org/23603>
- ECKERT D., BARON M., JÉGOU L. (2013). « Les villes et la science: apports de la spatialisation des données bibliométriques mondiales ». *M@ppemonde*, n° 110. <http://mappemonde.mgm.fr/num38/articles/art13201.html>
- GROSSETTI M., LOSEGO P., MILARD B. (1999). « La territorialisation comme contrepoint à l'internationalisation des activités scientifiques ». In LAREDO Ph., LERESCHE J.-Ph., WEBER K., *L'internationalisation des systèmes de recherche en action. Les cas français et suisse*, Lausanne: Presses Polytechniques et Universitaires Romandes (PPUR), p. 281-300. oai:halshs.archives-ouvertes.fr:halshs-00471192
- KRAEMER K., BOHR V. (2005). « The DNA Repair interest group: a global village ». *DNA Repair*, vol. 4, n° 4, p. 405–406. doi: 10.1016/j.dnarep.2005.01.002

- LAZEGA E., JOURDA M.-T., MOUNIER L., STOFER R. (2007). « Des poissons et des mares : l'analyse de réseaux multinationaux ». *Revue française de sociologie*, vol. 48, n° 1, p. 93-131. <http://www.cairn.info/revue-francaise-de-sociologie-2007-1-page-93.htm>
- LEYDESDORFF L., RAFOLS I. (2011). « Local emergence and global diffusion of research technologies : an exploration of patterns of network formation ». *Journal of the American Society for Information Science and Technology*, vol. 62, n° 5, p. 846–860. doi:10.1002/asi.21509
- MAISONOBE M., GIGLIA-MARI G., ECKERT D. (2013). « DNA Repair: A changing geography (1964-2008) ». *DNA Repair*, vol. 12, n° 7, p. 466-471. doi:10.1016/j.dnarep.2013.04.002
- MATTHIESSEN C. W., SCHWARZ A. W., FIND S. (2002) « The top-level global research system, 1997-99: centres, networks and nodality. An analysis based on bibliometric indicators ». *Urban Studies*, vol. 39, n° 5-6, p. 903-927. doi:10.1080/00420980220128372
- MULLINS N.-C. (1972). « The development of a scientific speciality : The Phage Group and the origins of Molecular Biology ». *Minerva*, vol. 10, n° 1, p. 51-82. doi:10.1007/BF01881390
- NEWMAN M.-E.-J. (2004). « Coauthorship networks and patterns of scientific collaboration ». *PNAS*, vol. 101, supplément n° 1, p. 5200–5205. doi:10.1073/pnas.0307545100
- POLANCO X. (1995). « Les sciences de l'information : bibliométrie, scientométrie, infométrie ». Solaris : Les Presses Universitaires de Rennes, n° 2.
- ROGERS E. M. (1962). *Diffusion of innovations*. New York: Free Press of Glencoe, 367 p.
- SOLLA PRICE D. (1963). *Little Science, Big Science*. New York: Columbia University Press, 118 p. ISBN: 0-231-08562-1
- POWELL W., WHITE D., KOPUT K., OWEN-SMITH J. (2005). « Network dynamics and field evolution: the growth of interorganizational collaboration in the Life Sciences ». *American Journal of Sociology*, vol. 110, n° 4, p. 1132–1205. doi:10.1086/421508
- SAINT-JULIEN Th. (2004). « Diffusion Spatiale ». *Hypergéométrie*, p. 1-4. [http://www.hypergeo.eu/IMG/\\_article\\_PDF/article\\_11.pdf](http://www.hypergeo.eu/IMG/_article_PDF/article_11.pdf)
- TER WAL A., BOSCHMA R. (2009). « Applying social network analysis in economic geography: framing some key analytic issues ». *The Annals of Regional Sciences*, vol. 43, n° 3, p. 739–756. doi:10.1007/s00168-008-0258-3
- VINCK D. (1999). « Les objets intermédiaires dans les réseaux de coopération scientifiques : contribution à la prise en compte des objets dans les dynamiques sociales ». *Revue française de sociologie*, p. 385-414.

## Notes

1. Il existe une corrélation entre le nombre de scientifiques rassemblés en un lieu et le volume de publications qui y est produit (Grossetti, 1999). Quant à l'intensité des relations entre les lieux de science, elle est fonction de leur volume de publications respectif (Matthiessen, 2002).
2. « The problem area can best be understood as a temporary unit which deals with special problems and then dissolves after one or several decades when the problems have either been solved or been determined to be unsolvable. »
3. Nous remercions tout particulièrement Giuseppina Giglia-Mari (CNRS, Toulouse) pour son aide et ses conseils.

4. Pour chaque période, il analyse la composition et l'organisation sociale du groupe. Lors de l'émergence, les membres sont peu nombreux et ne sont pas forcément en contact les uns avec les autres. Plus les opportunités de recherche se développent, plus le groupe attire de participants. Les relations se densifient jusqu'à ce que se forme un cluster bien identifié. Les membres reconnaissent leur appartenance à ce groupe et contribuent à son institutionnalisation, à sa légitimation en organisant des rencontres, des formations et en créant une ou des revue(s) spécialisée(s). Petit à petit, le groupe ne grandit plus, il est arrivé à saturation (en termes de potentiel scientifique) mais cherche à se maintenir socialement. Mullins situe son analyse au niveau des individus sans questionner l'effet de leurs localisations géographiques sur le collectif scientifique et son évolution.
5. En appliquant l'algorithme de visualisation Fuchterman-Reingold, les nœuds liés apparaissent proches les uns des autres et le rendu ne tient pas compte de la localisation géographique des acteurs du réseau (pour une discussion sur le choix de l'algorithme, voir, par exemple, Beauguitte, 2011).
6. Les liens réflexifs peuvent être importants en analyse multiniveau (Ducruet *et al.*, 2011) mais l'unité de base de la présente analyse de réseau est l'agglomération. Les collaborations locales sont exclues de l'analyse : le niveau local n'est pas le niveau d'organisation spatial de la recherche qui est mis en question dans le présent article.
7. L'indice de centralité d'intermédiation — ou « Betweenness » — d'un nœud  $i$  est le nombre de plus courts chemins du graphe passant par  $i$  sur l'ensemble des plus courts chemins du graphe.
8. Près de 80 % des 110 nouvelles agglomérations se trouvent dans seulement 10 pays (dans l'ordre : États-Unis, Japon, Italie, Royaume-Uni, Allemagne, Chine, Espagne, Canada, Finlande, France).
9. Pologne, Lettonie, Hongrie, Bulgarie, Slovaquie, Chine, Israël, Norvège, Belgique, Portugal et Singapour.
10. La communauté de la réparation de l'ADN est également très concentrée dans la triade. Cette organisation remonte aux années 1960 : voir le récent article publié dans la revue DNA Repair à ce sujet (Maisonobe *et al.*, 2013).
11. Six dans la première période et 20 dans la seconde n'ont jamais publié un seul article en collaboration avec une autre agglomération.
12. #Leiden, #Berkeley, #Kyoto, #Strasbourg, #Paris, #Washington, #New-York, #Baltimore, #Durham, #Tokyo, #Utrecht, #Brighton, #London, #Pavie, #Dallas, #Boston, #Toronto, #Galveston, #Aarhus, #Houston, #Livermore, #Los-Angeles, #Amsterdam, #Sao-Paulo, #Chicago, #Genève, #Kanazawa et #Zurich. Même si ce sont ces agglomérations qui sont les plus concernées par les effets d'agrégation des données, ces effets ne suffisent pas à expliquer ce résultat.
13. Les huit premiers producteurs : États-Unis, Pays-Bas, France, Japon, Royaume-Uni, Canada, Italie et Suisse. Les cinq nouveaux pays autonomes : L'Allemagne, le Brésil, l'Espagne, la Chine et la Finlande.
14. de 30 à 34 % en nombre et de 33 % à 48 % en valeur pour les liens supérieurs ou égaux à deux.
15. Aux États-Unis, cette part passe de 53 % (la plus élevée avant la part japonaise de 41 %) à 35 % (devenant inférieure à la part japonaise de 47 %).
16. La centralité de proximité ou « closeness centrality » : il s'agit de l'inverse de l'indice de Shimmel. Il se calcule pour un nœud donné  $i$  à partir de la distance de ce nœud  $i$  à tous les autres nœuds du graphe.
17. La valeur de lien minimale dans l'île est plus grande que la valeur de lien maximale au voisinage de l'île.
18. Aborder la question de recherche suppose de se connecter à ses instigateurs, Rotterdam en particulier, selon une logique d'attachement préférentiel : c'est ceux qui ont déjà le plus de liens qui continuent d'en recevoir le plus.
19. C'est à dire l'entretien de relations avec des semblables et des proches en qui l'on peut avoir confiance, cela se traduit en l'occurrence par une intensification des liens intra-nationaux.

## Adresse de l'auteur

Marion Maisonobe, UMR LISST, Université Toulouse 2-Le Mirail, 5 allées Antonio-Machado, 31058 Toulouse Cedex 9. Courriel : marion.maisonobe@univ-tlse2.fr