

Les graphes de cooccurrences technologiques pour l'analyse de l'innovation

Technological co-occurrence graphs for innovation analysis

Cécile Fauconnet¹

¹ Institut de Recherche Stratégique de l'Ecole Militaire (IRSEM), France, cecile.fauconnet@irsem.fr

RÉSUMÉ. Les graphes de cooccurrences technologiques en économie de l'innovation se construisent à l'aide des données de brevets, en y extrayant les classes technologiques associées aux inventions par les experts des offices de la propriété intellectuelle. Il y a cooccurrence lorsque qu'une même invention est allouée à plusieurs technologies. Cette dernière décennie a donné lieu à l'élaboration de nombreuses méthodes originales de traitement des données de brevets via ces graphes de cooccurrences. L'objectif de cet article est de faire un recensement et d'illustrer les questions de l'économie de l'innovation traitées à l'aide de graphes de cooccurrences technologiques.

ABSTRACT. Technological co-occurrence graphs in innovation economics are built using patent data, extracting the technological classes associated with inventions by experts from intellectual property offices. Co-occurrence occurs when the same invention is allocated to several technologies. The last decade has seen the development of many original methods for processing patent data via these co-occurrence graphs. The objective of this article is to identify and illustrate the issues of the innovation economy addressed by technological co-occurrence graphs.

MOTS-CLÉS. théorie des graphes, brevets, cooccurrences technologiques.

KEYWORDS. graph theory, patent, technological co-occurrence.

1. Introduction

La mobilisation des graphes pour l'étude des données de cooccurrences technologiques est de plus en plus courante pour l'étude de l'innovation. Cette dernière décennie a donné lieu à l'élaboration de nombreuses méthodes originales de traitement des données de brevets à travers l'extraction des classes technologiques associées aux inventions. L'objectif de cet article est de faire un recensement de ces méthodes et de leurs cas d'application, avec des exemples issus de cette littérature émergente et des données de la base COR&DIP [JRC 15].

L'usage des graphes pour l'étude de l'innovation trouve sa source dans une idée clé de [SCH 42] [USH 54] selon laquelle les innovations résultent de combinaisons nouvelles et non éprouvées d'éléments existants. La connaissance est ainsi caractérisée par sa nature cumulative. Le processus de recherche identifie et absorbe des connaissances pouvant être combinées pour générer de nouvelles connaissances [FLE 01] [SOR 06]. Ainsi, les connaissances sont le fondement de l'innovation et leurs combinaisons ne relèvent pas d'un processus aléatoire, elles sont réalisées de manière cohérente. Selon [PEN 59], cette accumulation et intégration de connaissances ne peut être réalisée efficacement que par l'ajout de connaissances connexes. Ce processus cumulatif et combinatoire détermine l'innovation en ce sens qu'il crée un « structure d'opportunité » pour l'innovation. Cette structure peut être visualisée à travers le concept de base de connaissances, définie par [HEN 90] comme étant le résultat de deux éléments distincts : les briques de connaissances et l'architecture des connaissances.

L'utilisation des graphes est ainsi particulièrement adaptée à l'étude de l'innovation en représentant les bases de connaissances où les connaissances sont les nœuds et leurs connexions les liens. De nombreux travaux en économie de l'innovation exploitent cette cohérence entre l'analyse du processus d'innovation et l'outils qu'est le graphe. Cependant, les recherches sur ces questions se distinguent par la diversité des données de cooccurrences mobilisées et par la multitude d'indicateurs développés à partir des graphes. L'analyse des réseaux permet de définir la structure de la base de connaissances à travers les notions de « degré », de « force », de « densité », de « proximité » (*closeness*), de « liaison »

(relatedness), ou encore de « cohérence ». Ces approches permettent de questionner le processus d'innovation autant sous l'angle de la base de connaissances d'un agent que dans une perspective de compréhension d'un secteur industriel ou encore de la diffusion technologique.

Cet article vise à présenter et discuter différentes applications de la théorie des graphes appliquée aux matrices de cooccurrences technologiques. Afin de construire cette représentation matricielle des bases de connaissances, les informations de classifications technologiques sont extraites des brevets. Le brevet est utilisé par les économistes comme un *proxy* de l'innovation technologique et apporte divers avantages, lorsqu'il est mobilisé pour s'intéresser au processus d'invention. Premièrement, les brevets représentent le résultat du processus inventif, et plus particulièrement des inventions dont on s'attend à ce qu'elles aient un impact commercial. Ils constituent une dimension particulièrement compétitive du changement technologique. Deuxièmement, l'obtention d'une protection par brevet prend du temps et coûte cher. Les demandes sont ainsi réalisées pour des inventions qui, en moyenne, devraient procurer des avantages compensant ces coûts. Troisièmement, les brevets sont classés par classe technologique, fournissant des informations non seulement sur le rythme de l'activité inventive, mais aussi sur son orientation. Enfin, les statistiques sur les brevets sont disponibles en grand nombre et pour de très longues séries chronologiques. En effet, le système des brevets est l'une des plus anciennes institutions des sociétés de marché [KAU 89] et il est conçu pour promouvoir et diffuser l'innovation.

Les données de brevets permettent ainsi aux économistes d'extraire les informations liées à la combinaison de connaissances technologiques. La Classification Internationale des Brevets (CIB)¹, « créée par l'Arrangement de Strasbourg de 1971, est un système hiérarchique de symboles indépendants de la langue pour le classement des brevets et des modèles d'utilité selon les différents domaines technologiques auxquels ils appartiennent »². Elle est conçue pour permettre un classement des brevets uniforme à l'échelon international. Elle se découpe en plusieurs niveaux d'agrégation exclusifs, de l'IPC 1-digit à l'IPC 7-digit. Le niveau 4-digit, le plus fréquemment utilisé par les économistes, est composé de 641 classes technologiques. Ces classes technologiques permettent de construire les graphes de cooccurrences. L'idée sous-jacente est que l'occurrence de deux technologies distinctes au sein d'un même brevet peut être modélisée comme une liaison entre ces deux classes technologiques.

Afin d'illustrer la diversité des recherches en économie de l'innovation mobilisant les graphes de cooccurrences technologiques, nous présentons trois perspectives de recherche distinctes. La section 2 présente l'utilisation des graphes pour la modélisation des innovations dans un domaine technologique donné. Dans la section 3 est développée l'apport des graphes à la définition et à la compréhension des bases de connaissances des entreprises à travers la notion de cohérence. Enfin, la section 4 discute des différentes possibilités d'approche de la diffusion des connaissances à l'aide les graphes de cooccurrences technologiques.

2. Domaines technologiques et graphes

Le renseignement de classes technologiques au sein des brevets par les experts des offices de la propriété intellectuelle facilite l'analyse de l'innovation à l'échelle des domaines technologiques. Dans la littérature, la définition des périmètres de ces domaines fait l'objet de deux approches principales. La première consiste en une construction du domaine technologique *ex post* par des experts technologiques à la manière des travaux de l'Observatoire des Sciences et Techniques (OST) ou de l'INPI, et la seconde se fonde sur une analyse automatisée du contenu textuel des brevets. Quelle que soit la définition choisie, la représentation sous forme de graphes des domaines technologiques permet de comprendre l'état de l'art dans un espace de connaissances donné ainsi que son évolution. Les

¹ International Patent Classification (IPC).

² <https://www.wipo.int/classifications/ipc/fr>

graphes de cooccurrences technologiques permettent de tracer l'environnement de connaissances dans lequel un domaine technologique se développe. Plus précisément, cette approche, à l'aide d'indicateurs de liaisons issues de la théorie de graphes, permet d'identifier les combinaisons de connaissances contribuant à l'innovation dans un domaine donné. Cette section s'attache à présenter cette approche, en statique et en dynamique, des domaines technologiques.

2.1. Approche statique : cartographie

Ainsi, une première approche de l'innovation technologique à l'aide de graphes de cooccurrences est la réalisation d'un état de l'art complémentaire à la cartographie des brevets présentées par Caillaud et Sternberger (ce numéro). [MEU 18] proposent une méthode générique pour produire ces panoramas et l'appliquent au cas particulier des innovations liées au microbiote intestinal. Cette méthode permet d'identifier des *clusters* technologiques autour d'un périmètre technologique donné à l'aide de mesure de liaisons entre les classes technologiques. Cette caractérisation d'un domaine technologique à l'aide des graphes de cooccurrences se réalise en plusieurs étapes : *i.* identification du périmètre technologique ; *ii.* mesure de la liaison technologique de ce périmètre avec le reste du paysage technologique ; *iii.* panorama des *clusters* technologiques participant à l'innovation du domaine technologique étudié. Afin de comprendre la manière dont cette méthode contribue à notre compréhension de l'innovation, nous présentons le cas d'application réalisé par [MEU 18].

Premièrement, [MEU 18] procèdent à une sélection des classes technologiques afin de tracer le périmètre du « système technologique d'innovation (STI) microbiote ». Ils utilisent leur périmètre à l'aide de recherches sémantiques dans la base de données brevets ORBIT (avec les entrées « microbiote », « microbiome » et « microflore ») et dans la Classification Internationale des Brevets (CIB) à son niveau le plus fin. Ces recherches sémantiques conduisent à identifier des « groupes » technologiques pertinents (tableau 1). Au-delà du traitement des troubles de l'appareil digestif (code A61P-001) et autres fonctions relatives à la santé humaine (code A61P-031), d'autres champs apparaissent, dans la préparation même des traitements, ou encore dans les cosmétiques.

Codes CIB	Intitulés
A61K-009	Préparations médicinales caractérisées par un aspect particulier
A61K-031	Préparations médicinales contenant des ingrédients actifs organiques
A61K-035	Préparations médicinales contenant des substances ou leurs produits de réaction de constitution non déterminée
A61K-045	Préparations médicinales contenant des ingrédients actifs non prévus dans les groupes
A61P-001	Médicaments pour le traitement des troubles du tractus alimentaire ou de l'appareil digestif
A61P-031	Agents anti-infectieux, c. à d. antibiotiques, antiseptiques, chimio thérapeutiques
A23L-001	Aliments ou produits alimentaires ; leur traitement, non couvert par d'autres classes
C12N-001	Micro-organismes, p.ex. protozoaires ; Compositions les contenant
A61K-008	Cosmétiques ou préparations similaires pour la toilette
C12Q-001	Procédés de mesure, de recherche ou d'analyse faisant intervenir des enzymes, des acides nucléiques ou des micro-organisme
A61K-038	Préparations médicinales contenant des peptides

Tableau 1. Codes CIB et intitulés des groupes technologiques relatifs au microbiote

Deuxièmement, les auteurs utilisent une mesure de la *relatedness* (liaison) générale [BRY 09] à partir des matrices de cooccurrences technologiques. Pour mesurer la *relatedness*, il faut extraire l'ensemble des codes « familles »³ de brevets allouant au moins l'une des 11 technologies du cœur

³ Une famille de brevets englobe tous les documents ayant exactement la même priorité ou plusieurs priorités en commun. En bref, elle permet de ne pas compter plusieurs fois une même invention ayant été déposée dans plusieurs offices. L'ensemble des illustrations au sein de cet article sont réalisée selon la définition INPADOC d'une famille de brevets. Une famille de brevets INPADOC englobe tous les documents couvrant une même technologie. Le contenu technique couvert par la demande est similaire, © 2021 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

microbiote (tableau 1) à l'échelle mondiale. Dans la base de données de brevets ORBIT, les auteurs identifient 12 885 familles de brevets déposés entre 2000 et 2018, 1 701 technologies y étant référencées (i.e. 1 690 en association avec les 11 du cœur microbiote). Ces 1 701 technologies concernées constituent l'environnement au sein duquel les inventions microbiotiques sont développées.

Clusters	# technologies	Descriptif
1	6	Traitement des troubles de l'appareil digestif
2	1	Produits alimentaires
3	1	Culture et conservation des microorganismes
4	1	Cosmétiques
5	1	Peptides antimicrobiens
6	1	Analyse médicale utilisant les microorganismes

Tableau 2. Clusters technologiques du microbiote

Troisièmement, cet ensemble de classes technologiques est agrégé en *clusters* à l'aide d'une technique de classification ascendante hiérarchique. 250 *clusters* technologiques sont recensés. Les 11 classes qui constituent le cœur technologique de l'innovation dans le microbiote se concentrent dans 6 *clusters*. Un *cluster* dominant se détache largement puisque qu'il concentre 6 des technologies centrales (tableau 2).

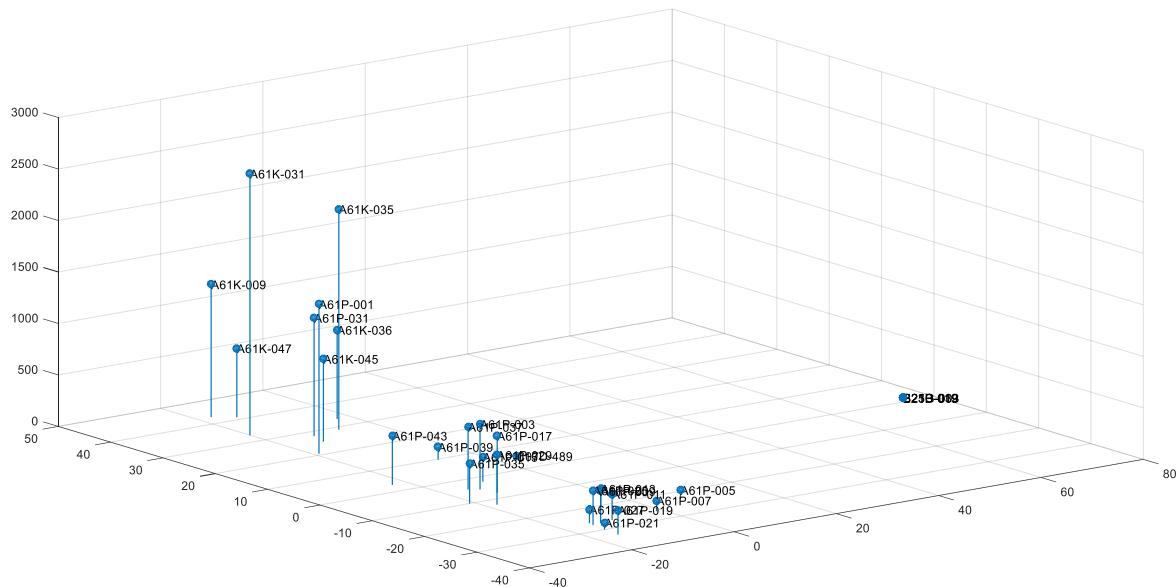


Figure 1. Composition du cluster « Traitement des troubles digestifs »

La figure 1 représente le *cluster* Traitement des troubles digestif, qui concentre les technologies A61K-031, A61K-035, A61K-009, A61P-031, A61P-001 et A61K-045 (voir tableau 1). La densité des technologies (1 701) et des *clusters* (250) fait qu'il est plus aisé d'interpréter ces derniers en les représentant un par un plutôt que dans leur ensemble. La hauteur des projections des technologies reflète l'intensité des inventions les concernant. Ainsi, environ 2 500 familles de brevets allouées à A61K-031 sont déposées entre 2000 et 2018.

mais pas nécessairement identique. Les membres d'une famille de brevets élargis auront au moins une priorité en commun avec au moins un autre membre, directement ou indirectement.

A cette approche purement technologique peuvent être ajouter des échelles d'analyses en étudiant par exemple la place d'un brevet dans un *cluster* technologique ainsi que la contribution d'une entreprise à l'innovation dans un espace donné de la connaissance. A titre d'illustration, [MEU 18] présentent la contribution du groupe L'Oréal au *cluster* Cosmétique identifié précédemment. La lecture de la figure 2 apporte deux informations importantes :

- l'intensité des références aux technologies du cluster, qui correspond au nombre de revendications faites par l'acteur (somme des hauteurs en rouge dans le graphique) ;
- la couverture de l'acteur dans le *cluster*, qui correspond au rapport entre le nombre de technologies qu'il produit dans ce *cluster* (en rouge) et le nombre total de technologies que ce *cluster* contient.

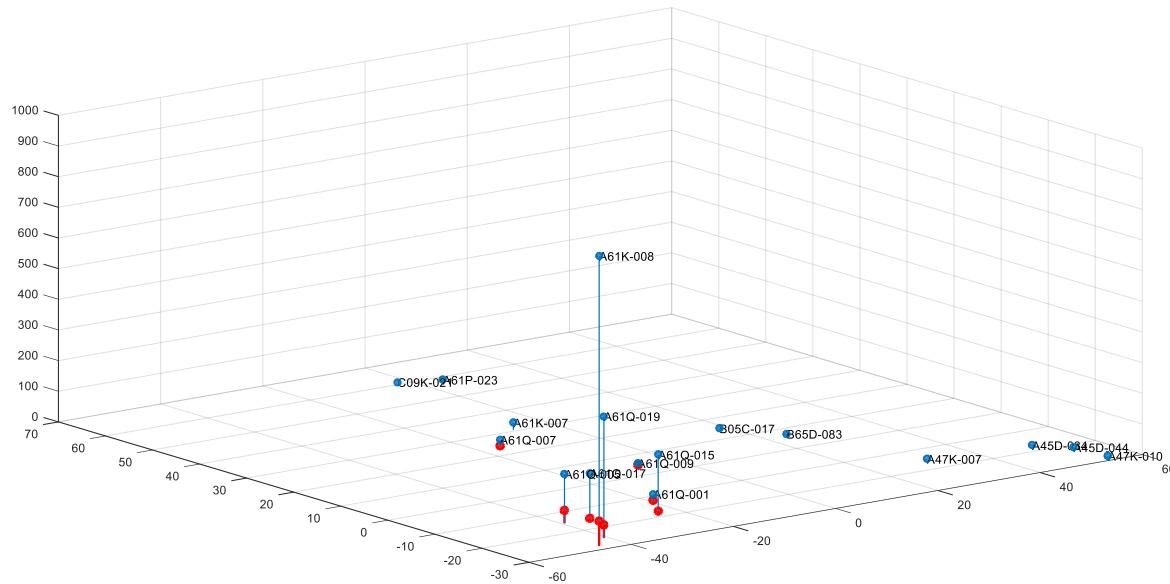


Figure 2. Contribution de L'Oréal (en rouge) au cluster Cosmétiques (en bleu)

[MEU 18] en déduisent des indicateurs d'analyse de la politique R&D des groupes au niveau de chaque *cluster* microbiotique. Les graphes de cooccurrences peuvent produire en définitive un outil d'aide à la prise de décision industrielle.

2.2. Approche dynamique

En complément à cette approche statique des domaines technologiques, d'autres auteurs proposent des analyses de l'évolution de ces domaines sur longue période. [KRA 11] utilisent des données de cooccurrences de technologies allouées dans des brevets pour représenter la dynamique sur longue période d'un domaine technologique particulier : celui des biotechnologies. Ils en définissent classiquement le périmètre en termes de codes de la nomenclature de la Classification Internationale des Brevets (CIB ; cf. tableau 3). Ils représentent des liens forts entre les technologies de ce domaine et avec des technologies connexes à la manière des figures 3 et 4.

Les graphes de cooccurrences représentent les bases de connaissances spécifiques associées à ce domaine technologique. Nous actualisons les recherches de [KRA 11] en mobilisant des données de brevets plus récentes tout en reproduisant la démarche des auteurs. Nous constatons une densification des combinaisons autour des biotechnologies, une hausse du nombre des technologies connexes, et le maintien d'un découpage interne à ce domaine entre ce qui relève des équipements, des applications médicales et des applications végétales. En d'autres termes, la base de connaissances des biotechnologies s'enrichit mais maintient un degré élevé de spécialisation.

Codes CIB	Intitulés
A01H	Nouveautés végétales ou procédés pour leur obtention ...
A61K	Préparations à usage médical, dentaire ou pour la toilette
C02F	Traitements de l'eau, des eaux résiduaires, des eaux ou boues d'égout
C07G	Composés de constitution indéterminée
C07K	Peptides
C12M	Appareillage pour l'enzymologie ou la microbiologie
C12N	Micro-organismes ou enzymes ...
C12P	Procédés de fermentation ...
C12Q	Procédés de mesure ou de test faisant intervenir des enzymes ...
G01N	Recherche ou analyse des matériaux par détermination de leurs propriétés chimiques ou physiques

Tableau 3. Domaine des biotechnologies

Au-delà de la visualisation de ces tendances, comme dans les figures 3 et 4, des outils de l'analyse des réseaux sociaux [WAS 94] sont utilisés pour donner corps à ce constat. La densité des connexions, par exemple, se mesure simplement en comptant le nombre de cellules non vides de la matrice d'adjacence du graphe sur le nombre total de cellules de cette matrice moins le nombre de nœuds du graphe. La centralité d'un nœud du graphe, *i.e.* son importance relative, dépend chez ces auteurs de la distance moyenne (le nombre moyen de liens) qui lui permet de se connecter à tous les autres nœuds du réseau (centralité dite de proximité), et ils utilisent également d'autres mesures de centralité issue de cette littérature [FRE 78]. Ce qui est assez intéressant à noter, c'est que dans tous les indicateurs qui viennent d'être décrits, seules les informations sur le nombre des connexions des nœuds / technologies sont mobilisées.

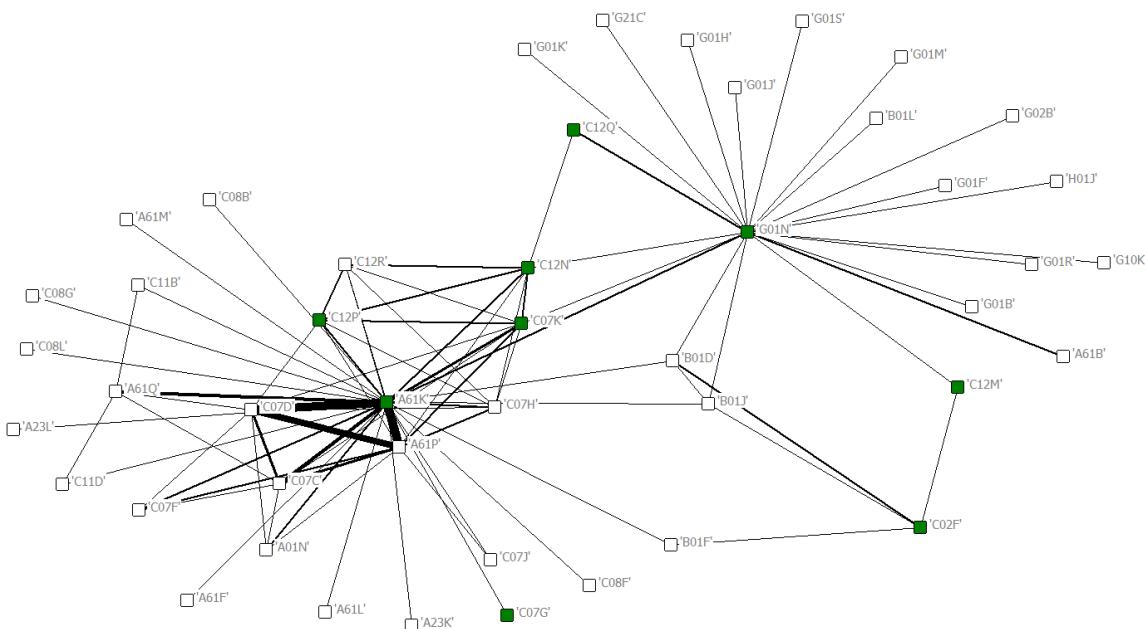


Figure 3. Structure des liens technologiques forts dans le domaine des biotechnologies – 1980-1982

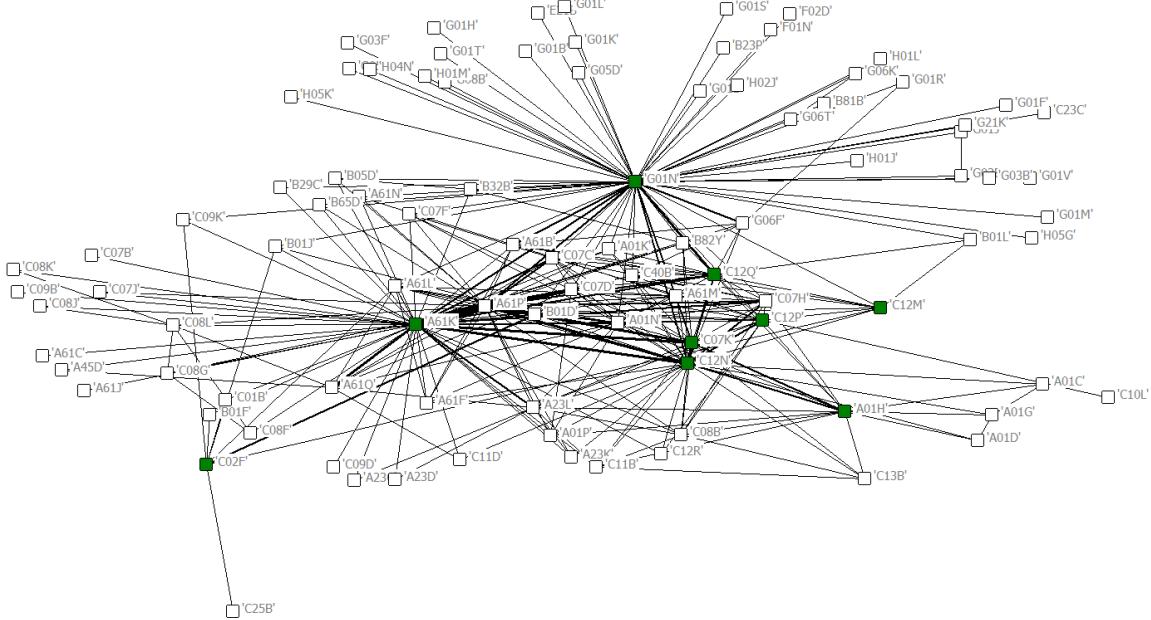


Figure 4. Structure des liens technologiques forts dans le domaine des biotechnologies – 2010-2012

Lecture. Les nœuds du graphe représentent les briques technologiques (codes CIB), les liens entre ces nœuds des architectures de connaissances (combinaisons technologiques associées dans les brevets déposés dans le domaine des biotechnologies). Les nœuds en vert sont les technologies biotech ([KRA 11] ; cf. tableau 1), ceux en blanc sont les autres technologies auxquelles les premières s'associent « fortement ». Une connexion entre deux technologies / nœuds est représentée dans ces graphes uniquement si elle est « forte ». Il faut au moins cent relations entre deux technologies au sein des brevets pour qu'une liaison entre elles soit considérée comme forte. L'épaisseur des liaisons indique cette force sur une échelle allant de 1 (proche de 100 cooccurrences) à 10 (proche de 1000 cooccurrences). Dépôts à l'OEB, l'USPTO et les offices de l'UE15 (« familles INPADOC »), périodes 1980-1982 et 2010-2012. Les intitulés des codes CIB sont disponibles sur www.wipo.int/classifications/ ipc/fr/

Cette approche est cependant très descriptive. Le niveau d'analyse n'est ni celui de l'entreprise, ni celui de l'industrie mais d'un agglomérat de technologies, agglomérat qui n'est pas un objet d'analyse habituel en économie industrielle. Avec des bases de données plus récentes, il devient possible d'ajouter les unités « entreprise » et « industrie » pour l'étude des dynamiques d'innovation – par exemple COR&DIP, qui est depuis 2015 une base de données mise à disposition par l'OCDE sur un échantillon de groupes très actifs en R&D à l'échelle mondiale. Nous proposons dans la section suivante une application qui prenne en compte l'entreprise et ses brevets dans l'étude des cooccurrences technologiques.

3. Base de connaissances des entreprises et graphes

La cohérence aide à mieux comprendre les stratégies de diversification des entreprises à travers l'approche de la *relatedness* [ANS 57]. Celle-ci vise à mesurer les liens existants entre des domaines industriels ou technologiques en évaluant le degré de connexions entre les connaissances. Plus les briques de connaissances sont reliées entre elles, plus la base de connaissances de l'entreprise est cohérente. Le premier traitement systématique de la cohérence a été fourni par les travaux de [TEE 94]. Ils développent une méthode pour mesurer la cohérence de l'entreprise en fonction de ses *outputs*. Les auteurs considèrent qu'une entreprise n'est pas une collection d'activités sans rapport entre elles, mais qu'elle doit se diversifier de façon cohérente. Ils soutiennent que les entreprises cohérentes ont plus de chances de réussir économiquement que les entreprises « incohérentes ». Cela s'explique par l'existence d'économies d'échelle liées à la diversification des activités dans des domaines connexes. L'étude de la cohérence technologique d'une entreprise s'effectue en deux temps :

- la construction du graphe de cooccurrences, indépendamment des frontières des entreprises, que [ELY 15] appellent « paysage de cohérence » ;
- une analyse statistique rapportant les caractéristiques de l’entreprise aux propriétés du paysage de cohérence, pour mesurer son degré de cohérence.

Ces indicateurs de cohérence technologique de la base de connaissances des entreprises sont largement mobilisés dans la littérature afin de comprendre l’innovation et la performance économique des entreprises. [TEE 94] se concentrent sur l’articulation entre les activités productives au sein des entreprises. Ils trouvent qu’à mesure que les entreprises manufacturières américaines se diversifient, elles maintiennent un niveau constant de cohérence entre les activités voisines. Cette cohérence aide à mieux comprendre la manière dont les connaissances sont structurées dans l’entreprise en distinguant les combinaisons de connaissances communément utilisées (forte cohérence) ou si elles sont le fruit d’associations technologiques inhabituelles (faible cohérence). De nombreux raffinement de cette cohérence ont été proposés notamment pour adresser empiriquement la question du dilemme entre l’exploration et l’exploitation mis en avant par [MAR 91]. Ce dernier montre qu’il existe différentes façons de tirer parti des connaissances existantes, soit pour exploiter des combinaisons technologiques connues, soit pour explorer de nouvelles possibilités. Il y a, d’une part, l’exploitation des synergies déjà connues entre les ressources des entreprises afin d’accroître les liens entre elles ([RUM 74], [MON 79]). D’autre part, il y a l’exploration de nouveaux liens entre les ressources de l’entreprise dont le potentiel synergique n’est pas encore identifié et qui peuvent conduire à une combinaison inhabituelle de ressources [NES 05], pouvant générer des coûts organisationnels prohibitifs [REE 91].

De nombreux auteurs mobilisent cet indicateur de cohérence technologique afin de mieux comprendre les performances des entreprises. Par exemple, [NES 05] mobilisent la cohérence technologique afin d’expliquer la performance à l’innovation des entreprises pharmaceutiques américaines dans les années 1990. Ils mettent en évidence que les flux de connaissances sont nécessaires à l’innovation mais ne suffisent pas à garantir sa qualité. De même, [NES 06] examinent la relation entre les caractéristiques de la base de connaissances des entreprises en termes de capital de connaissances et d’intégration des connaissances et la valeur sur le marché boursier de 84 entreprises actives dans la biotechnologie au cours des années 1990. Ils montrent que l’intégration de connaissances devient un déterminant de plus en plus important de la performance de marché. Selon les auteurs, ce résultat illustre l’intégration croissante de la biotechnologie dans de nombreuses applications industrielles. Plus récemment, [FAU 18] s’intéressent à la cohérence technologique des entreprises de l’aérospatiale en France entre 2012 et 2014 et mettent en évidence l’importance de la relation entre diversification technologique et l’exploration de combinaisons de connaissances originales. Dans la même dynamique, [FAU 19] utilisent les indicateurs de cohérence afin d’expliquer la performance économique des entreprises sur un marché particulièrement demandeur d’innovation technologique, celui de la défense. Leurs résultats mettent en avant l’importance de l’exploration pour le chiffre d’affaires de défense des entreprises.

Afin d’illustrer cette cohérence de la base de connaissances, la figure 5 présente une projection du groupe Nexter sur le paysage de cohérence européen (données OEB sur les années 2010 à 2012, base de données COR&DIP de l’OCDE). Quelles technologies ce groupe combine-t-il pour produire ses inventions ? Les technologies (codes CIB à 4-digit) sont les nœuds du réseau et les liens entre ces nœuds représentent des cooccurrences technologiques de trois manières différentes :

- soit le groupe « exploite » des liaisons technologiques assez communément établies dans le paysage de cohérence, et elles apparaissent en rouge dans le graphe ;
- soit le groupe établit des liaisons entre technologies qui sont relativement peu présentes dans ce paysage. Dans ce cas, on peut dire que ces liaisons « différencient » l’entreprise dans son activité de R&D. Ces liens apparaissent en gris ;

– soit le groupe décide de ne pas exploiter des connexions technologiques qui sont par ailleurs largement combinées (liens verts). Dans ce cas, le groupe adopte un comportement de « distinction ».

Que se passerait-il en termes de cohérence d'organisation de la R&D si Nexter décidait d'associer des chercheurs pour exploiter un « lien vert », ou d'abandonner quelques « liens gris » qui font sa spécificité ? Comment l'expert R&D Nexter interpréterait-il cette projection sur le paysage de cohérence : comme des liens à construire car mal perçus, ou au contraire comme des choix stratégiques raisonnés ?

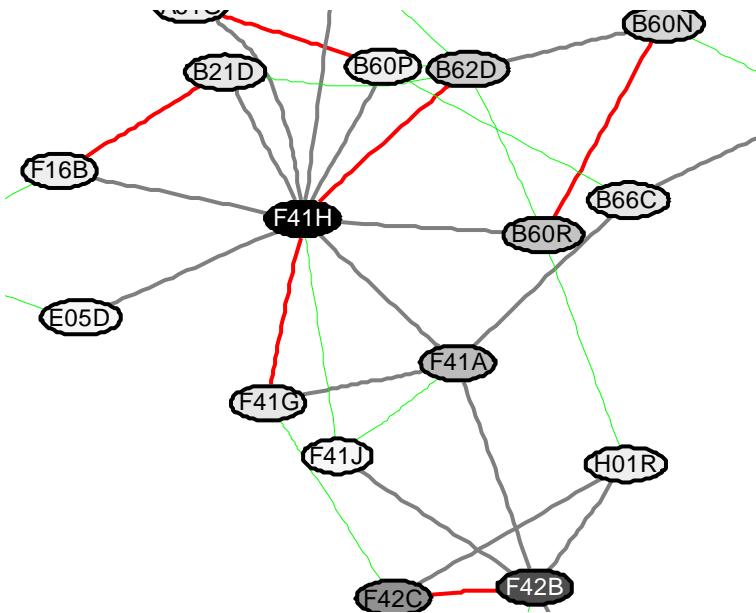


Figure 5. Extrait d'une projection du portefeuille de brevets du groupe Nexter sur le paysage de cohérence du marché européen des technologies (2010-2012)

Ainsi, l'analyse de la cohérence issues des graphes de cooccurrences est à la fois un outil pour la recherche en science économique afin de mieux appréhender les processus d'innovation mais peut aussi contribuer à la prise de décision stratégique pour l'organisation de la R&D des entreprises.

4. Diffusion des connaissances et graphes

En complément à ces approches à l'échelle de domaines technologiques et des entreprises, les graphes issus des matrices de cooccurrences technologiques permettent d'adresser des questionnements plus généraux, notamment à travers la notion de diffusion technologique dans l'ensemble de la société. En science économique, la diffusion technologique est essentiellement abordée sous l'angle des usages [GER 00]. Elle ne se résume cependant pas à cela. S'il fallait raisonner schématiquement, trois approches de la diffusion technologique se distinguent :

– « 1, puis 10, puis 100 ». La première approche met l'accent sur les phénomènes d'amplification, d'intensification temporelle de l'adoption, à périmètres technologiques inchangés. Les choix du marché impulsent la dynamique, et révèlent des phénomènes comme les rendements croissants d'adoption et la dépendance au sentier par exemple (voir [DOS 10], pour une synthèse). C'est la compétition technologique qui est au cœur de l'analyse ;

– « A, puis AB, puis ACDE ». La deuxième approche met l'accent sur les combinatoires non cumulatives dans la production de connaissances. Les périmètres technologiques sont mouvants, des « trajectoires » se dessinent, des « paradigmes » se succèdent (voir [GKO 16], pour une synthèse). Plus

une technologie est combinée à d'autres en *t*, plus elle est importante à cette date dans ce processus de production d'inventions ;

– « A, puis AB, puis ABCDE ». La troisième approche relève de combinatoires cumulatives dans la production de connaissances nouvelles. Comme dans l'approche précédente, les périmètres technologiques sont mouvants, et les thèmes abordés ont trait aux « états de l'art », à la *pervasiveness* des technologies, à leur ubiquité. Plus une technologie est, dans son histoire, combinée à d'autres, plus elle est importante aujourd'hui dans le processus de production d'invention global.

Les deux dernières approches exploitent une base empirique commune : la cooccurrence des technologies associées aux inventions, que l'on identifie dans les brevets à partir de l'usage généralisé des nomenclatures technologiques. Les graphes qui en sont déduits sont étudiés en utilisant des outils de l'analyse des réseaux [WAS 94] au niveau des « degrés » des nœuds, c'est-à-dire du nombre total d'autres technologies auxquelles ils se lient pour produire des inventions.

[GKO 16] proposent un modèle structural de diffusion illustratif de la deuxième approche (combinatoires non cumulatives). Meunier (ce numéro) propose une alternative prenant en compte la nature cumulative dans la production de connaissances (les manières anciennes de produire des inventions ne se perdent pas définitivement). Dans les deux cas, les auteurs montrent que la dynamique de diffusion peut correctement être appréciée par des courbes en S, résultat classique dans cette littérature.

Dans le cas des combinatoires non cumulatives, [GKO 16] posent que la valeur limite de diffusion se fixe au maximum observé à leur date d'étude. Il devient alors possible de prédire le temps nécessaire pour atteindre la date de maturité d'une technologie pour n'importe laquelle d'entre elles avec pour objectif d'identifier celles ayant un fort potentiel de maturation *à court terme*. L'approche non cumulative, en fixant le potentiel de diffusion des technologies, pose selon nous une condition trop forte, qui est celle d'attribuer une ubiquité réalisable importante à *toutes* les technologies, c'est-à-dire de poser qu'à terme toute technologie sera capable de se combiner *à beaucoup* d'autres, ce qui nous semble aller à l'encontre des résultats empiriques de l'économie de l'innovation [GER 00]. Dans le cas des combinatoires cumulatives, Meunier endogénéise la limite de diffusion, sans pour autant contraindre la date à laquelle la maturation technologique se produit. Il estime donc un paramètre supplémentaire, avec pour objectif d'identifier les technologies à fort potentiel absolu *à long terme*.

Ces deux approches structurales de la diffusion technologique ne sont pas antinomiques. Elles répondent en fait à deux enjeux différents. Dans l'approche non cumulative, il s'agit de cibler les technologies qui vont prochainement basculer dans la maturité, c'est-à-dire possédant un potentiel de diffusion à court terme élevé. Dans l'approche cumulative, il s'agit d'identifier les technologies qui possèdent le potentiel de développement le plus élevé en absolu, que ce potentiel se réalise à court ou à plus long terme. Il est facile d'envisager un graphe non cumulatif en date *t* comme une composante du graphe cumulatif à cette même date. L'approche cumulative intègre donc l'approche non cumulative. Autrement dit, on ne perd pas d'information ou de capacité d'analyse en faisant le choix du cumulatif.

L'approche cumulative possède en plus des potentialités de développement importantes, notamment sur deux points qui nous semblent essentiels. Premièrement, elle permet de prendre en compte des intensités de connexions, ce que ne fait aucun modèle structural de diffusion technologique aujourd'hui. Dans les modèles de [GKO 16] et de Meunier, seules des informations de degré sur les graphes de cooccurrences technologiques sont mobilisées. L'information apportée par les « forces » dans les modèles cumulatifs permettraient de produire des analyses sur l'émergence de nouvelles trajectoires technologiques (montée relative en intensité dans le temps de certaines combinaisons), sur le déclin d'autres, sur les changements de paradigmes (bouleversement des structures de connexions en termes d'existence et / ou d'intensité sur un temps court). Ce sont des objets d'analyse qui se situent au cœur même de l'approche non cumulative.

Deuxièmement, l'approche cumulative permet de prédire les connexions nouvelles à court terme des technologies en fonction des structures passées et présentes de leurs connexions. Comparons deux structures de connexions en degrés et en forces de deux technologies non encore connectées entre elles. Imaginons que nous soyons capables de mesurer une distance entre ces structures. On peut poser l'hypothèse que plus les structures sont proches l'une de l'autre, plus la probabilité future de connexion entre ces technologies est élevée. A notre connaissance, la littérature qui permettrait d'étayer cette hypothèse n'existe pas en économie de l'innovation. Par contre, dans d'autres domaines de recherche en économie où les techniques structurales sont utilisées, cette hypothèse a été émise et vérifiée par l'intermédiaire d'indices de distance spécifiques. Il serait donc possible d'appliquer ces indices à un nouveau terrain d'investigation, celui de l'innovation technologique. Comme domaine d'inspiration premier, nous pensons à l'économie internationale, plus précisément à l'étude des structures commerciales par produit et par pays qui est menée par [HID 11]. L'idée de cet auteur est de montrer qu'il existe une logique d'insertion des pays en développement au commerce mondial qui passe par une diversification productive dans des secteurs dont les structures d'exportation sont proches de celles qu'il exploite déjà. La diversification se fait ici à la marge, à partir de compétences que l'on maîtrise en grande partie. La distance entre structures de connexions est un *proxy* d'une similarité entre compétences. C'est cette similarité de compétences qui pourrait supporter l'hypothèse posée plus haut.

5. Conclusion

Cet article passe en revue les différentes questions de l'économie de l'innovation adressées à l'aide de graphes de cooccurrences technologiques. De la question des dynamiques d'innovation dans un domaine technologique, au potentiel de diffusion des connaissances, en passant par l'analyse du processus d'innovation des entreprises, l'usage des techniques structurales contribue à une large part de notre compréhension de la place des connaissances et de l'innovation dans notre société. La cooccurrence technologique repérée dans les brevets à travers la nomenclature CIB constitue la voie privilégiée pour associer la capacité inventive à la combinaison de connaissances et en comprendre les transformations. L'intérêt des méthodes présentées dans cet article est qu'elles conduisent à aborder ensemble les différentes sources de transformation.

A l'échelle d'une entreprise, ces transformations peuvent se faire en interne (par l'exploration de nouvelles combinaisons que l'entreprise n'a, de son côté, jamais testées) ou bien en externe (par prise de contrôle capitalistique ou en développant des partenariats en R&D). En effet, d'un point de vue formel, des acquisitions ou cessions d'actifs ou des relations contractuelles longues avec des partenaires en R&D ne font que changer le périmètre de l'analyse. L'analyse structurale permet d'étudier ces questions en reliant la dimension entreprise à celle des domaines technologiques. Reprenons le cas de l'article de [KRA 11] : quelle que soit l'organisation du réseau de connaissances des biotechnologies, elle tourne autour d'un cœur parfaitement identifié et, dans le temps, les connexions entre les éléments de ce cœur ainsi que les technologies externes qui s'y rattachent changent. Ce sont ces changements au niveau technologique qui conduisent à envisager leurs impacts sectoriels, à la manière de [BER 15], et c'est la richesse de la méthode structurale de pouvoir établir empiriquement un lien vertical entre la technologie et le marché.

Après 10 ans de ces travaux, nous ne sommes qu'aux balbutiements de cette compréhension. Les enjeux actuels et futurs de la recherche sur ces thématiques sont multiples, et incluent notamment la problématique de traitement des données sémantiques contenues dans les brevets et de leurs représentations sous forme de graphes.

Bibliographie

- [ANS 57] ANSOFF H.I., « Strategies for diversification », *Harvard Business Review*, p. 113-124, 1957.

- [ANS 65] ANSOFF H.I., *Corporate strategy: an analytic approach to business policy for growth and expansion*, McGraw-Hill, 1965.
- [BER 15] BERGEK A., HEKKERT M., JACOBSSON S., MARKARD J., SANDÉN B., TRUFFER D., « Technological innovation systems in contexts: conceptualizing contextual structures and interaction dynamics », *Environmental Innovation and Societal Transitions*, vol. 16, p. 51-64, 2015.
- [BRY 09] BRYCE D.J., WINTER S.G., « A general interindustry relatedness index », *Management Science*, vol. 55, p. 1570-1585, 2009.
- [DOS 10] DOSI G., NELSON R.R., « Chapter 3 - Technical change and industrial dynamics as evolutionary processes », *Handbook of the Economics of Innovation*, vol. 1, 2010.
- [ELY 15] EL YOUNSI H., LEBERT D., MEUNIER F.-X., ZYLA C., « Exploration, exploitation et cohérence technologique », *Economie Appliquée*, vol. 68 (3), p. 187-204, 2015.
- [FAU 18] FAUCONNET C., LEBERT D., MOURA S., ZYLA C., « Les entreprises aérospatiales de défense : quelles cohérences technologiques de leur R&D ? », *Technologie et Innovation*, vol. 3(4), 2018.
- [FAU 19] FAUCONNET C., MEUNIER F.-X., « Major R&D firms: how defence technological innovation contributes to their arm sales? », 45th annual conference of Eastearn Economic Association, New York, 2019.
- [FLE 01] FLEMING L., « Recombinant uncertainty in technological search », *Management Science*, vol. 47, p. 117-132, 2001.
- [FRE 78] FREEMAN L.C., « Centrality in social networks: conceptual clarification », *Social Networks*, vol. 1, p. 215-239, 1978.
- [GER 00] GEROSKI P.A., « Models of technology diffusion », *Research Policy*, vol. 29, p. 603-625, 2000.
- [GKO 16] GKOTSIDIS P., VEZZANI A., « Technological diffusion as a recombinant process », *JRC Working Papers on Corporate R&D and Innovation*, n° 07/2016, 2016.
- [HEN 90] HENDERSON R.M., CLARK K.B., « Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms », *Administrative Science Quarterly*, vol. 35(1), p. 9-30, 1990.
- [HID 11] HIDALGO C.A., « Discovering Southern and East Africa's industrial opportunities », *Economic Policy Paper Series, The German Marshall Fund of the United States*, 2011.
- [JRC 15] JRC/OECD., « 2015 COR&DIP v.1 DATABASE », 2015.
- [KAU 89] KAUFER E., *The economics of the patent system*, Harwood Academic Publishers, 1989.
- [KRA 11] KRAFFT J., QUATRARO F., SAVIOTTI P.P., « The knowledge-base evolution in biotechnology: a social network analysis », *Economics of Innovation and New Technology*, vol. 20, p. 445-475, 2011.
- [MAR 91] MARCH J.G., « Exploration and exploitation in organizational learning », *Organization Science*, vol. 2, p. 71-87, 1991.
- [MEU 18] MEUNIER F.-X., LEBERT D., « Explorer l'activité d'innovation technologique autour d'une fonction : le cas du traitement des troubles digestifs par le microbiote », halshs-02024381, 2018.
- [MON 79] MONTGOMERY C.A., *Diversification, market structure, and firm performance: an extension of Rumelt's model*, Purdue University, 1979.
- [NES 05] NESTA L., SAVIOTTI P.P., « Coherence of the knowledge base and the firm's innovative performance: evidence from the US pharmaceutical industry ». *The Journal of Industrial Economics*, vol. 53, p. 123-142, 2005.
- [NES 06] NESTA L., SAVIOTTI P.P., « Firm knowledge and market value in biotechnology », *Industrial and Corporate Change*, vol. 15, p. 625-652, 2006.
- [PEN 59] PENROSE E., *The theory of the growth of the firm*, OUP, Oxford, 1959.
- [REE 91] REED M., « Bimodality in diversification: an efficiency and effectiveness rationale », *Managerial and Decision Economics*, vol. 12, p. 57-66, 1991.
- [RUM 74] RUMELT R.P., *Strategy, structure, and economic performance*, Division of Research, Graduate School of Business Administration, Harvard University, 1974.
- [SCH 42] SCHUMPETER J.A. *Capitalism, socialism and democracy*, Routledge, 1942.
- [SOR 06] SORENSEN O., RIVKIN J.W., FLEMING L., « Complexity, networks and knowledge flow », *Research Policy*, vol. 35, p. 994-1017, 2006.

- [TEE 94] TEECE D.J., RUMELT R.P., DOSI G., WINTER S.G., « Understanding corporate coherence: theory and evidence », *Journal of Economic Behavior & Organization*, vol. 23, p. 1-30, 1994.
- [USH 54] USHER A.P., *A history of mechanical inventions*, Courier Corporation, 1954.
- [WAS 94] WASSERMAN S., FAUST K., *Social network analysis: methods and applications*, Cambridge University Press, 1994.