

# Segmentation sociale d'inventeurs : le cas de l'industrie de la nanotechnologie au Canada

AHMAD BARIRANI, CATHERINE BEAUDRY, BRUNO AGARD

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

C.P. 6079, succ. Centre-ville, Montréal (Québec), H3C 3A7, Canada

[ahmad.barirani@polymtl.ca](mailto:ahmad.barirani@polymtl.ca); [catherine.beaudry@polymtl.ca](mailto:catherine.beaudry@polymtl.ca); [bruno.agard@polymtl.ca](mailto:bruno.agard@polymtl.ca)

---

**Résumé** – À l'ère de la globalisation, il est de plus en plus important pour les pays avancés d'améliorer leur capacité d'innovation afin de conserver leur avance technologique vis-à-vis des pays en émergence. Pour y arriver, certains modes de pensées et pratiques relatives à la gestion de l'innovation doivent être revus afin de permettre un meilleur rendement des innovateurs. L'analyse des réseaux d'inventeurs a récemment reçu beaucoup d'attention dans les milieux académiques. Ces techniques permettent d'observer les interactions sociales des inventeurs dans leur milieu de pratique et de voir comment le capital social de chaque inventeur a un impact sur son rendement en innovation. Cet article propose une nouvelle méthode d'analyse des réseaux d'inventeurs. Notre méthodologie consiste d'abord à segmenter les inventeurs suivant leur capital social pour ensuite mesurer le rendement moyen de chaque segment trouvé. Cette méthode sera appliquée au cas des inventeurs canadiens de brevets en nanotechnologie. Nous pourrions ainsi identifier les types de profils sociaux qui sont les plus productifs au sein de la communauté d'inventeurs canadiens en nanotechnologies.

**Abstract** – In an era of globalization, it becomes increasingly important for advanced countries to increase their innovation capabilities in order to stay ahead technologically against emerging countries. To achieve this objective, certain lines of thought and practices regarding innovation management need to be readjusted in order to achieve a better performance. Inventor network analysis has recently received a lot of attention in academia. These techniques are used to observe inventors' social interactions within their environment of practice and see how different social interaction patterns have an impact on inventors' performance. This article proposes a new method for analysing inventor networks. Our methodology first consists in clustering inventors according to their social capital and then in measuring the average performance for each segment. This method will be applied to the case of Canadian nanotechnology patent inventors. By doing so, we will be able to identify the most productive types of social profiles within the Canadian nanotechnology inventors' community.

**Mots clés** – gestion de l'innovation, réseaux d'inventions, réseaux sociaux, exploration de texte, segmentation.

**Keywords** – innovation management, invention networks, social network, text mining, clustering.

---

## 1 INTRODUCTION

À l'ère de la globalisation, un nouveau paysage économique marqué par l'hyper-compétitivité est en train de se former. Dans un premier temps, de grands pas dans le domaine de l'ingénierie industrielle et de la logistique ont offert la possibilité aux entreprises manufacturières de déplacer leurs activités de production dans des pays où la main d'œuvre était meilleur marché. Ainsi, après une longue période de rationalisation industrielle et de montée en compétence des fournisseurs, nous sommes face à une situation où ces pays jadis en voie de développement sont aujourd'hui des chefs de file dans plusieurs domaines technologiques de pointe. Cette nouvelle réalité a une incidence importante sur l'organisation industrielle des pays avancés qui sont forcés de développer des produits et services innovants à un rythme plus important pour assurer leur développement économique.

Il devient ainsi important de revoir certaines pratiques en matière de gestion de l'innovation et de mettre en place des politiques qui favorisent l'innovation au sein des communautés scientifique et technologique des pays avancés. Dans cet article, nous proposons d'atteindre cet objectif en analysant les

réseaux d'inventeurs canadiens en nanotechnologie afin de trouver les profils sociaux qui mènent à de meilleurs rendements en termes de production de brevets dans ce secteur.

Dans cet article, nous proposons une méthodologie qui permettra de monter un réseau social d'inventeurs, de faire la segmentation des inventeurs sur la base de leur capital social et de trouver les profils sociaux qui sont les plus productifs sur le nombre de brevets produits. Cette méthodologie sera appliquée au cas des réseaux d'inventeurs canadiens en nanotechnologie pour lequel nous expliquerons les résultats.

## 2 REVUE DE LITTÉRATURE

### 2.1 Gestion de l'innovation

La littérature concernant la gestion de l'innovation peut être divisée en deux grandes écoles. La première voit la capacité d'innovation comme une caractéristique intrinsèque de tout agent économique [Amin et Cohendet, 2004]. Il s'agit là de l'approche de *gestion stratégique* où les gestionnaires sont les seuls *entités pensantes* de la firme. Ceux-ci se basent sur les

compétences clés de la firme pour prendre des décisions sur les directions stratégiques à prendre. Le rôle des gestionnaires revient alors à mettre en place des structures qui permettent à l'entreprise de protéger et de renforcer ses compétences. Par conséquent, les politiques nationales qui se rapportent à la favorisation de l'innovation sont limitées à l'identification et le support financier de quelques innovateurs potentiels ou importants.

De l'autre côté, la deuxième école de pensée se préoccupe plutôt de ce que l'environnement a à offrir aux agents économiques [Amin et Cohendet, 2004]. Ici, la capacité d'innovation d'un agent économique est fonction des connaissances que les autres agents ont à lui offrir. Il s'agit ici de reconnaître l'importance des externalités réseaux dans le processus de création et de diffusion de l'innovation. Ainsi, la principale qualité d'un innovateur est dans sa capacité d'absorption du savoir qui lui vient de l'extérieur [Cohen et Levinthal, 1990]. Naturellement, les politiques qui découlent de ce mode de réflexion tentent de mettre en place des mécanismes qui favorisent la collaboration et le transfert de connaissances entre les agents innovants. Récemment, cette voie théorique a reçu beaucoup d'attention dans le milieu académique [Burt, 2004 ; Adler et Kwon, 2002 ; Inkpen et Tsang, 2005]. Dans de nombreux projets de recherche [Watts et Strogatz, 1998 ; Newman, 2003 ; Cowan et Jonard, 2003], les techniques d'analyse de réseaux sociaux sont utilisées pour étudier l'impact des interactions entre inventeurs sur le rendement de ceux-ci en innovation.

## 2.2 Mesure de performance

Un bureau de brevet octroie un brevet à un cessionnaire lorsqu'une invention déposée par cette dernière est utile, non-évidente et est une nouveauté. Ces caractéristiques font en sorte que les brevets peuvent être utilisés comme indicateurs de performance en innovation [Griliches, 1990]. Toutefois, les brevets ne sont pas représentatifs de toute la production innovante d'un pays. En effet, il existe d'autres méthodes de protection de la propriété intellectuelle.

De plus, le brevet n'est pas toujours synonyme d'innovation. De nombreux brevets ne se manifestent jamais sous forme de produits ou ne génèrent pas de bénéfices pour les firmes cessionnaires. [Griliches, 1990] propose l'utilisation du nombre de citations reçues par un brevet pour mesurer sa valeur économique, hypothèse qui est supportée dans une étude empirique menée par [Harhoff *et al.*, 1999].

## 2.3 Analyse de réseaux sociaux

[Wasserman et Faust, 1994] fournissent une description exhaustive des éléments requis lors de l'analyse de réseaux sociaux. Un réseau est défini par une paire d'ensembles  $G = \{P, E\}$  où  $P$  est un ensemble de  $N$  nœuds  $P_1, P_2, \dots, P_n$  et  $E$  est un ensemble de liens qui connectent deux éléments de  $P$ .

Chaque nœud a un *degré de distribution* qui est défini par le nombre de liens qui le relie aux autres nœuds du réseau. Le nombre de liens réseaux qui séparent deux nœuds  $i$  et  $j$  s'appelle la *distance géodésique*. Le *plus court chemin* est la distance géodésique la plus petite qui sépare deux nœuds  $i$  et  $j$ . Nous appelons *proximité (closeness centrality)* la distance géodésique moyenne qui sépare un nœud  $i$  de tous les autres nœuds du réseau. Le *degré d'intermédiation (betweenness centrality)* d'un nœud  $i$  est défini par

$$C_B(i) = \sum_{j \neq k \neq i} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$$

où  $\sigma_{jk}$  est le nombre de plus courts chemins entre les nœuds  $j$  et

$k$  et  $\sigma_{jk}(i)$  est le nombre de plus courts chemins entre  $j$  et  $k$  qui passent par  $i$ . Le *coefficient de groupement (clustering coefficient)*  $C_i$  d'un nœud  $i$  est défini par

$$C_i = \frac{2E_i}{K_i(K_i - 1)}$$

où  $E_i$  représente le nombre de liens entre les  $k_i$  nœuds qui sont liés au nœud  $i$ . Cette métrique montre à quel point les nœuds reliés à un nœud  $i$  sont reliés entre eux. Un ensemble de nœuds qui sont tous liés entre eux font partie d'une *clique*.

La figure 1 montre un exemple de réseau social. Les nœuds 1, 2 et 3 font partie d'une clique. Les nœuds 1 et 3 ont un coefficient de groupement de 1,0, alors que le nœud 2 a en a un de 1/3. Les degrés d'intermédiation pour les nœuds 2 et 4 sont de 8 et de 9,5 respectivement puisque les nœuds 1 et 3 doivent *passer* par ceux-ci pour atteindre les nœuds 5, 6 ou 7. Ces derniers sont des nœuds plutôt périphériques, ils se retrouvent moins souvent sur un plus court chemin et ont donc un degré d'intermédiation plus petit.

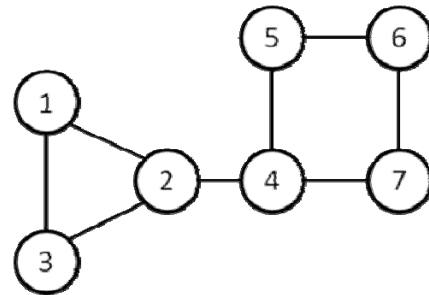


Figure 1. Réseau social comprenant 7 nœuds et 8 liens.

### 2.3.1 Réseaux aléatoires et petits-mondes

Dans un réseau composé de  $n$  nœuds, il y a une possibilité totale de  $n(n-1)/2$  liens entre les nœuds. Toutefois, les différents types de réseaux que nous trouvons dans la nature ne montrent pas autant de liens entre les nœuds. Un *réseau aléatoire* est un réseau pour lequel les liens sont choisis aléatoirement parmi tous les nœuds possible entre les  $n$  nœuds. En changeant la probabilité  $p$  avec laquelle deux nœuds peuvent être connectés, on peut obtenir une gamme de réseaux aléatoires qui se situe entre le réseau régulier ( $p = 0$ ) et le réseau complet ( $p = 1$ ).

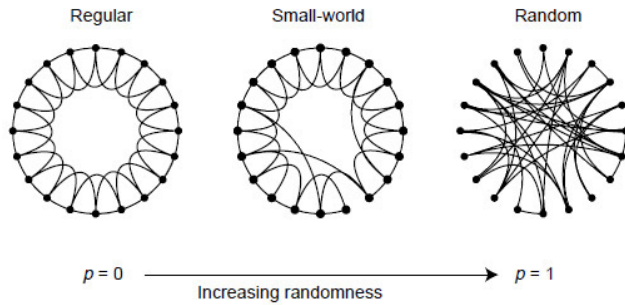
La plupart des réseaux retrouvés dans la nature sont de type *petit-monde (small-world)* [Watts et Strogatz, 1998], i.e. le degré de proximité moyen des nœuds du réseau est petit et le coefficient de groupement est plus grand que dans un réseau aléatoire. La figure 2 montre un schéma du petit-monde se situant entre un réseau régulier et un réseau entièrement aléatoire.

Les petits-mondes sont reconnus pour favoriser le transfert de ce qui peut être transmis d'un nœud à l'autre dans le réseau. Ainsi, un réseau d'inventeur qui aurait les caractéristiques d'un petit-monde favoriserait la diffusion plus rapide des connaissances entre des inventeurs du réseau [Cowan et Jonard, 2003]. De plus, dans un réseau d'inventeurs de type petit-monde, on observe que le niveau de connaissance agrégé des inventeurs est plus grand que dans les réseaux aléatoires.

### 2.3.2 Capital social et performance

Les métriques telles que le degré de centralité, de groupement et d'intermédiation peuvent être vues comme un *capital social* que chaque individu possède dans un réseaux [Nahapiet et

Ghoshal, 1998]. De nombreuses études montrent qu'il y a un lien entre le capital social d'un individu et ses performances. Par exemple, [Burt, 2004] montre que des gestionnaires qui bénéficient d'une position d'intermédiaires dans un réseau de contact de gestionnaires ont plus souvent de bonnes idées et sont mieux récompensés que les autres gestionnaires du réseau. [Cantner et Graf, 2006] montrent que les innovateurs qui ont une position centrale au sein de leur réseau innovent à un rythme plus important et jouissent de cette position centrale pendant une plus longue période.



**Figure 2. Le petit-monde se situe entre un réseau régulier et un réseau aléatoire [Watts et Strogatz, 1998].**

Il est à noter que la centralité et l'intermédiation sont des caractéristiques sociales qui sont plus souvent associées à des milieux qui sont technologiquement turbulents. Ces caractéristiques sont moins prisées lorsque des industries ont atteint une maturité et que les activités innovantes se concentrent sur l'amélioration des procédés en place et l'introduction d'innovations incrémentales.

#### 2.4 Segmentation

Les techniques de segmentation consistent à regrouper un ensemble d'observations de façon à ce que les éléments similaires se retrouvent dans le même segment [Berry et Linoff, 2004]. Il s'agit là d'une technique d'apprentissage non-supervisée.

Il existe plusieurs types de méthodes de segmentation. Toutes les méthodes basées sur la similarité ont en commun le fait de requérir une mesure de distance entre deux éléments. La distance euclidienne, la distance Manhattan et le coefficient de corrélation de Pearson en sont des exemples. La segmentation vise à ce que la distance inter-segment soit maximisée et que la distance intra-segment soit minimisée [Manning *et al.*, 2008]. Les sections suivantes font état de l'état de l'art des trois méthodes de segmentations les plus répandues.

##### 2.4.1 Segmentation hiérarchique

La segmentation hiérarchique classe les éléments d'observations sous une structure en arbre après un certain nombre d'itérations [Berry et Linoff, 2004]. La segmentation peut se faire soit par agglomération (de bas en haut : algorithmes CHA, CURE) ou par division (de haut en bas : algorithmes DIANE, BIRCH). Dans les méthodes par agglomération, chaque segment constitue initialement un segment. À chaque itération, les segments les plus proches sont fusionnés pour former de nouveaux segments élargis. Le processus se termine lorsqu'il n'y a plus qu'un seul segment. Les méthodes par division, à l'inverse, partent avec un seul segment qui contient tous les éléments. À chaque itération, les segments sont divisés de manière à ce que les éléments d'un segment soient les plus distants possible des éléments des

autres segments. Le processus se termine lorsque tous les segments ne contiennent qu'un seul élément.

##### 2.4.2 Segmentation par partitionnement

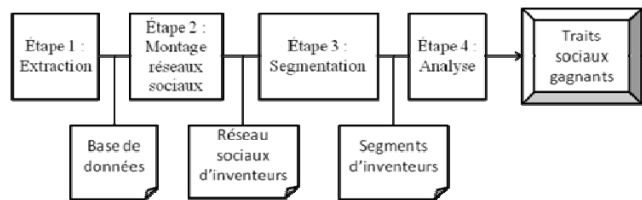
La segmentation par partitionnement (ex : *k*-means) regroupe les données selon un nombre fixe (*k*) de segments. Ce nombre de segments peut être prédéfini par l'utilisateur ou trouvé après quelques tentatives avec différents essais. Le processus de partitionnement commence en assignant un élément à un segment. Cet élément deviendra le noyau du segment. Ensuite, chaque élément restant est assigné à un segment selon la distance de l'élément avec le noyau du segment. Il y a alors une itération où un nouveau noyau est choisi pour chaque segment. Les éléments sont alors réassignés à des segments toujours selon leur distance avec le noyau [Berry et Linoff, 2004]. Le processus s'arrête après un nombre d'itérations prédéfini ou lorsque le processus converge vers un optimum local.

##### 2.4.3 Segmentation par densité

La segmentation par densité diffère de la segmentation par partitionnement du fait qu'elle n'impose pas un nombre prédéfini de segments. Les segments sont plutôt définis par des *régions* qui comprennent un nombre suffisant d'éléments proches [Ester *et al.*, 1996 ; Ankerst *et al.*, 1999]. Toutefois, ces méthodes prennent en paramètres d'entrée le seuil de similarité  $\epsilon$  (la distance maximum au-delà de laquelle deux éléments ne peuvent être considérés comme étant proches) et le nombre d'éléments minimum pour former un segment. Le processus de segmentation débute en choisissant un élément qui fera partie d'un segment. Tous les éléments qui seront à l'intérieur de  $\epsilon$  seront affectés au segment. De même, tous les éléments qui sont à une distance  $\epsilon$  de tout élément du segment y sont affectés. Lorsqu'il n'y a plus d'éléments à considérer pour ce segment, un élément qui n'est pas assigné à un segment est choisi pour faire parti d'un autre segment. De la même façon, tous les éléments qui sont à une distance  $\epsilon$  d'un des éléments de ce segment en feront partie. Le processus se termine lorsque tous les éléments sont assignés à un segment ou lorsqu'aucun des éléments non-assignés n'est en deçà d'une distance  $\epsilon$  d'un segment.

### 3 MÉTHODOLOGIE

Notre méthode consiste à utiliser une technique de segmentation pour explorer les réseaux sociaux et trouver des profils sociaux d'inventeurs qui pourraient donner lieu à de meilleurs rendements d'innovation. La méthodologie proposée compte quatre étapes telles qu'illustrées dans la figure 3.



**Figure 3. Méthodologie**

#### 3.1 Étapes successives

**Étape 1 :** Dans un premier temps, nous procédons à l'extraction et au nettoyage d'un échantillon de données provenant de bases de données de brevets de l'United States Patent and Trademark Office (USPTO) [USPTO, 2009]. Nous extrayons de cette base de données : les titre, résumé, dates

d'application et d'obtention, nombre de réclamations, références, citations, nom, ville et pays des inventeurs ainsi que nom, ville et pays des cessionnaires pour chaque brevet d'intérêt. Les brevets sont sélectionnés en se basant sur une liste de mots clés. Nous utiliserons la méthode d'*extraction binaire* [Manning *et al.*, 2008], ce qui signifie que les brevets extraits contiendront au moins l'un des mots clés apparaissant dans la liste.

**Étape 2 :** En second temps, nous montons et étudions la structure des réseaux sociaux d'inventeurs canadiens en nanotechnologie. À cet effet, le réseau social des inventeurs sera un réseau dans lequel les nœuds sont des inventeurs provenant de notre échantillon d'inventeurs extrait à l'étape 1 et dont les liens sont définis par la relation de co-invention identifiée par la liste des inventeurs de chaque brevet.

Dans le cadre de la présente étude, nous définissons le capital social de chaque inventeur par son degré de distribution, d'intermédiation, de proximité et par le coefficient de groupement. L'objectif est de mesurer le capital social de chaque inventeur, de voir à quel point le réseau social d'inventeurs est fragmenté et à quel point il présente les caractéristiques d'un petit-monde.

**Étape 3 :** La troisième étape de ce projet consiste à effectuer une segmentation des inventeurs d'après le capital social de chacun. Pour ce faire, nous sélectionnons la composante réseau principale et nous utilisons les métriques de capital social de chaque inventeur comme attributs des éléments à segmenter. La plus importante composante est celle qui contient le plus grand nombre d'inventeurs. Travailler sur la plus importante composante permet dans un premier temps d'améliorer la précision de notre échantillon initial de brevets. En effet, notre méthode d'extraction basée sur les mots clés aura forcément un certain nombre de faux positifs. Dans un second temps, effectuer une segmentation sur un nombre plus grand d'inventeurs permet d'obtenir des résultats plus robustes.

**Étape 4 :** Finalement, une fois les segments obtenus, nous procédons à la caractérisation des segments en calculant la moyenne pour les métriques de capital social des inventeurs de chaque segment. Chaque segment représente un type de profil social dans lequel nous pouvons classer les inventeurs. Par exemple, un segment qui sera majoritairement composé d'inventeurs qui font davantage partie de cliques sera caractérisé comme étant un segment composé d'individus qui ont des liens forts. En effet, l'existence d'une clique indique qu'il y a une affinité suffisamment grande entre les membres de la clique. C'est cette affinité qui fait en sorte que les membres de la clique soient liés les uns aux autres [Granovetter, 1973]. Par la suite, nous calculons les métriques se rapportant à la performance moyenne de chaque segment. Les métriques utilisées seront le nombre moyen de brevets par inventeur, le nombre moyen de citations obtenues pour les brevets au cours des 5 années suivant leur date d'octroi.

### 3.2 Outils et techniques utilisés

Les brevets proviennent des bases de données de l'USPTO. Ce choix est motivé par l'importance que représente le marché américain sur le plan technologique et commercial. Pour effectuer notre recherche par mot clés et extraire les données, nous utilisons le logiciel PatentBot développé à l'École Polytechnique de Montréal. Ce logiciel exécute l'interface web de l'USPTO pour exécuter les requêtes et extraire les données relatifs aux brevets ciblés. De plus, nous avons eu accès à la base de données de l'OST pour obtenir l'information relative aux citations des brevets extraits via PatentBot.

Afin de monter ce réseau, nous allons utiliser le logiciel *open source* NodeXL [CodePlex, 2011] en fournissant l'ensemble des liens entre les inventeurs obtenus à partir des données USPTO. Ce logiciel permet de tracer le graphe social d'inventeurs et de calculer les métriques que nous allons utiliser durant l'étape de segmentation.

Nous utilisons le logiciel *open source* RapidMiner [Rapid-I, 2011] pour la segmentation des données. À cet effet, l'algorithme de segmentation DBSCAN est utilisé. Ce choix est motivé par le fait que nous ne voulons pas, *a priori*, imposer un nombre strict de profils sociaux dans lequel chaque inventeur devra se retrouver. Cette méthode permet de laisser libre court à toute la diversité qu'il peut y avoir dans la structure sociale des inventeurs.

## 4 CAS D'APPLICATION

### 4.1 Contexte

Nous avons choisi le domaine des nanotechnologies pour effectuer notre analyse. Ce choix se justifie par le fait que la nanotechnologie est un domaine technologique multidisciplinaire en émergence dans lequel les différents sous-domaines ne sont pas encore tous en interaction.

Les industries en émergence sont reconnues pour connaître initialement des phases turbulentes où beaucoup d'innovations sont introduites mais où le niveau de réussite de ces innovations est très variable. L'industrie de la nanotechnologie est donc riche en données qui permettent de faire une bonne discrimination entre les inventeurs qui y évoluent, ce qui en fait un exemple approprié pour y appliquer les techniques de segmentation.

Finalement, les nanotechnologies sont généralement reconnues comme étant l'un des domaines de pointe les plus importants sur le plan technologique dans les années à venir. Comme il a été discuté précédemment dans cet article, la capacité des pays avancés à assurer leur compétitivité dans ce domaine est cruciale pour leur avenir technologique. Le Canada ne fait pas exception, ce qui nous motive à regarder de plus près la structure du réseau technologique canadien dans ce domaine.

### 4.2 Hypothèses

Puisque l'industrie de la nanotechnologie peut être considérée comme étant jeune, en émergence et caractérisée par l'interdisciplinarité, nous pouvons émettre l'hypothèse que les métriques qui montrent qu'il y a une diversité dans le capital social des inventeurs sont de meilleurs indicateurs de performance que les métriques qui montrent qu'il y a une similarité dans le capital social des inventeurs. Nous nous attendons à ce que les segments d'inventeurs qui ont un degré de centralité plus petit et qui ont un coefficient de groupement plus élevés obtiennent moins de brevets et de citations par invention que les segments d'inventeurs qui ont un degré d'intermédiation et de centralité plus élevé.

Sur le plan de la structure du réseau social, nous nous attendons à ce que le réseau soit très fragmenté. Cette hypothèse est justifiée à prime abord par le fait que les réseaux d'inventions sont généralement fragmentés et ensuite par le fait que la nanotechnologie est un domaine jeune, pour lequel peu de standards et de technologies commercialisées, pouvant faire partie d'une plateforme technologique sur laquelle la communauté d'inventeurs convergerait ses activités, existent.

### 4.3 Extraction des données

Nous avons utilisé la liste de mots clés apparaissant au tableau 1. Ces mots clés ont été obtenus à partir des mots clés qui ont été utilisés dans au moins deux études bibliographiques

antérieures concernant les nanotechnologies [Alexcar *et al.*, 2007; Fitzgibbons et McNiven, 2006; Mogoutov et Kahane, 2007; Porter *et al.*, 2008; Schmoch *et al.*, 2003; Zitt et Bassecoulard, 2006]. Ces mots clés ont été utilisés pour extraire les brevets déposés de 1976 à 2009 en nanotechnologies dont le cessionnaire ou l'un des inventeurs est canadien. Ainsi, les brevets octroyés à des firmes non canadiennes mais dans lequel un inventeur canadien est nommé sont aussi considérés dans notre analyse.

La base de données de brevets extraits est ensuite nettoyée afin d'enlever les duplicatas, les entrées incomplètes ou mal formatées. Après le nettoyage, sur les 8 076 brevets du tableau 1, 5 811 ont été retenus pour l'étape suivante. Ces brevets ont été inventés par plus de 18 223 inventeurs, ce qui fait une moyenne de 2,26 inventeurs par brevet. Dans notre échantillon d'inventeurs, plus de 13 165 résident au Canada.

**Tableau 1. Mots clés nanotechnologies**

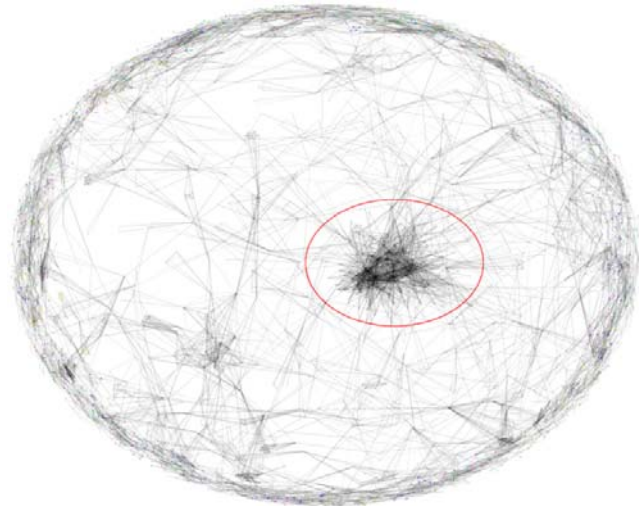
	Nombre de brevets extraits
nano*	4 568
atom* force microscop*	88
biosensor	231
mesoporous material*	31
molecular beam epitaxy	95
molecular switch	25
nems	9
polymer composite*	379
polymer dna	10
polymer rna	3
quantum	1287
scanning probe microscop*	16
self assem*	219
supramolecular chemistry	18
tunnel* microscop*	2
photonic*	969
scanning prob*	41
single electron*	85

#### 4.4 Réseaux sociaux

Les liens entre inventeurs se font par l'intermédiaire de brevets, peu importe qu'ils soient octroyés à des entreprises canadiennes ou non. Les inventeurs qui résident au Canada et qui participent à des inventions qui sont octroyées à des entreprises non-canadiennes peuvent quand même être reliés à des inventeurs qui résident au Canada mais qui brevettent pour des entreprises canadiennes. Aussi, dans le cas des multinationales, il peut arriver que des inventeurs canadiens travaillent conjointement avec des inventeurs qui sont à l'étranger sur quelques brevets. Notons aussi que notre échantillon contient des brevets sur une période de 33 ans (de 1976 à 2009) et le réseau social qui en résulte tient compte de tous les liens cumulés sur ces années. Puisque nous voulons trouver le lien entre capital social et performance en innovation, il semble judicieux de procéder de la même manière. En effet, un lien de co-invention implique un échange de connaissances entre les co-inventeurs. L'accumulation du capital social signifie alors une accumulation de connaissances. En tenant compte de l'ensemble des liens sur une longue période, nous mesurons aussi l'impact de l'accumulation des liens sociaux passés sur les performances futures.

Il y a au total 4 543 inventeurs qui sont socialement reliés à d'autres inventeurs. Ces inventeurs ont un total de 9 384 liens entre eux, ce qui fait une moyenne de 2,07 liens par inventeur.

Ainsi, sur les 13 165 inventeurs canadiens de notre échantillon, 8 622 (65%) sont des singletons, i.e. qu'ils n'ont aucun lien avec d'autres inventeurs. On peut alors supposer que le réseau social des inventeurs canadiens en nanotechnologie est très fragmenté et que beaucoup d'inventeurs n'ont de liens de co-invention qu'avec deux autres inventeurs. Ceci est en parti dû au fait que les singletons sont enlevés de l'échantillon des 4534 inventeurs connectés et que ceux-ci sont en moyenne co-inventeurs avec 2 autres inventeurs.



**Figure 4. Graphe du réseau social des inventeurs canadiens en nanotechnologie. Il faut noter que tous les nœuds ne sont pas interconnectés. La partie encerclée représente la principale clique de la composante principale.**

Toujours à l'aide de NodeXL, nous avons sélectionné la composante réseau principale qui comprend 945 inventeurs soit 20,8% de tous les inventeurs du réseau. Nous mesurons le coefficient de groupement, le degré de distribution, le degré de proximité et le degré d'intermédiation pour chaque inventeur du composant pour l'étape suivante de notre analyse. Les moyennes des résultats sont affichées au tableau 2.

**Tableau 2. Capital social moyen et écart-type du composant principal**

	Moyenne	Écart-type
Degré de distribution	7,07	8,85
Degré d'intermédiation	4 537	21 335
Coefficient de groupement	0,69	0,36
Degré de proximité	$1,0 \times 10^{-4}$	$2,0 \times 10^{-4}$

#### 4.5 Segmentation

Étant donnée la grande variation dans la magnitude des indicateurs de capital social (Tableau 2), nous avons dû normaliser en calculant les indices  $z$  pour les attributs de capital social de chaque inventeur de la composante principale. Chaque inventeur peut se distinguer des autres par la diversité dans son capital social. Par exemple, un inventeur peut faire partie d'une clique tout en ayant un degré de distribution et d'intermédiation élevé. D'autres joueront un rôle plutôt périphérique dans le réseau et auront un degré de distribution

plus petit que les autres. Puisque nous allons utiliser quatre attributs pour représenter le capital social de chaque inventeur, les possibilités de trouver plusieurs types de profils sociaux sont importantes. Puisque l'un des objectifs de cet article est de découvrir des profils sociaux, nous ne voulions pas imposer un nombre strict de segments dans lesquels les inventeurs seraient classés. Notre objectif est plutôt de faire un regroupement *naturel* des inventeurs qui ont un profil social suffisamment proche. Afin de pouvoir atteindre cet objectif, nous avons choisi d'adopter la méthode de segmentation par densité DBSCAN.

Toujours en lien avec notre objectif de faire une segmentation raffinée, nous avons choisi une valeur  $\varepsilon$  de 0,5 et un nombre minimum d'inventeurs par segments de 10. La valeur  $\varepsilon$  est justifiée par le fait que les attributs sociaux sont normalisés. En choisissant la valeur de 0,5, nous pouvons regrouper les inventeurs qui sont à  $\frac{1}{2}$  écart-type de distance, ce qui permet de faire un regroupement plus raffiné des inventeurs qui sont loin de la normale.

**Tableau 3. Segmentation de la composante principale des inventeurs canadiens en nanotechnologie**

	Nombre d'inventeurs	Nombre de brevets
Segment 1	100	990
Segment 2	89	137
Segment 3	730	1133
Segment 4	3	89
Segment 5	12	314
Segment 6	10	386

#### 4.6 Analyse des segments

Les Tableaux 4 et 5 montrent le capital social moyen pour les segments trouvés à l'étape précédente. En analysant les attributs sociaux de chaque segment, on voit qu'il y a des différences dans le profil social des inventeurs qui figurent dans chaque segment. Par exemple, le segment 2 est principalement composé d'inventeurs qui ont très peu de liens au sein du réseau. Le segment 3 regroupe aussi des inventeurs qui ont moins de liens mais qui ont un coefficient de groupement plus élevé que la moyenne. Ce segment est probablement composé d'inventeurs qui ont développé un petit nombre de brevets au sein d'un groupe fermé d'inventeurs. On peut déduire que les segments 2 et 3, regroupant 87% des inventeurs, sont principalement composés d'inventeurs périphériques d'après leur profil social.

**Tableau 4. Capital social moyen (segments 1, 2 et 3)**

	Segment		
	1	2	3
Degré de distribution	16,24	1,11	5,46
Degré d'intermédiation	37 141	97	688
Coefficient de groupement	0,34	0	0,84
Degré de proximité	$1,2 \times 10^{-4}$	$9,7 \times 10^{-5}$	$1,0 \times 10^{-4}$

Contrairement aux inventeurs du segment 3, ceux du segment 1 ont plus de liens, mais un coefficient de groupement plus petit. Ce segment se distingue d'ailleurs de tous les autres segments par le degré d'intermédiation de ses inventeurs. Si on compare le segment 1 avec les segments 4, 5 et 6, on voit que

ces derniers ont plus de liens réseau mais qu'ils ont une position d'intermédiaire moins importante. Cette différence au niveau du degré d'intermédiation s'explique par le fait que les inventeurs du segment 1 sont connectés en moyenne avec des inventeurs qui ont eux aussi un degré d'intermédiation élevé. Ainsi, même si les inventeurs du segment 1 ont moins de liens, ils ont un degré d'intermédiation plus important parce que les inventeurs auxquels ils sont connectés sont aussi d'importants intermédiaires. Il faut aussi noter que les segments 4, 5 et 6 sont beaucoup plus petits comparés aux autres segments, ce qui indiquerait que leur profil social soient plutôt l'exception. Puisque ces derniers ont un degré de centralité beaucoup plus grand que la grande majorité des inventeurs, on peut supposer que ces derniers sont des inventeurs étoiles. Ce concept s'approche de celui des *chercheurs étoiles* [Zucker et al., 1998] qui représentent un petit pourcentage des chercheurs et qui sont reconnus pour être beaucoup plus productifs que les autres. L'analyse des segments à la section suivante pourra apporter quelques éclaircissements à ce niveau, à savoir si un haut degré de distribution peut être associé à l'obtention d'un plus grand nombre de brevets et de citations par brevets.

**Tableau 5. Capital social moyen (segments 4, 5 et 6)**

	Segment		
	4	5	6
Degré de distribution	38,67	32,92	45,8
Degré d'intermédiation	2 410	1 483	3 330
Coefficient de groupement	0,32	0,34	0,28
Degré de proximité	$1,2 \times 10^{-4}$	$1,2 \times 10^{-4}$	$1,2 \times 10^{-4}$

#### 4.7 Analyse des performances

Comme on peut le voir sur le tableau 6, les segments 2 et 3 sont les moins performants en termes de nombre de brevets publiés par inventeur. En effet, il s'agit là d'un grand nombre d'inventeurs qui ont moins de 2 brevets en moyenne. Comme on peut prévoir, les inventeurs du segment 1 sont plus prolifiques en termes de brevets. Toutefois, les inventeurs des segments 5 et 6 sont exceptionnellement performants avec en moyenne 26,2 et 38,6 brevets par inventeur. Les performances de ces derniers en terme de brevets obtenus se rapproche plus du concept des inventeurs étoiles. Ces résultats sont très intéressants puisque nous nous attendions à ce que le segment 1 soit le plus performant puisque les inventeurs dans ce segment ont des positions d'intermédiaire plus importantes. Ainsi, nos résultats suggèrent que le degré de distribution semble être le trait social le plus important lorsque l'on considère le nombre de brevets obtenus comme mesure de performance.

**Tableau 6. Performance des inventeurs par segment**

	Nombre de brevets	Nombre de brevets moyen par inventeur
Segment 1	990	9,9
Segment 2	137	1,5
Segment 3	1 133	1,6
Segment 4	89	7
Segment 5	314	26,2
Segment 6	386	38,6

Pour ce qui est de la performance des brevets obtenus par les inventeurs, il semble y avoir aussi une importante variation entre les segments. Tel que montré dans le tableau 7, les brevets du segment 1 sont plus souvent cités que ceux du segment 2. Les inventeurs qui ont un haut degré de distribution et d'intermédiation reçoivent en moyenne 43% plus de citations que ceux qui sont plus périphériques.

Toutefois, on observe que les inventeurs du segment 3 produisent des brevets qui sont cités plus souvent que ceux de ce dernier. À première vue, ce résultat suggérerait que les inventeurs qui ont un degré de centralité plus bas que la moyenne et qui font partie d'une clique produisent des brevets qui sont d'une qualité aussi importante que ceux qui ont un haut degré de centralité et d'intermédiation. Par contre, il ne faut pas oublier que le fait qu'ils ont tous un coefficient de groupement égale à 1 peut aussi signifier qu'ils soient co-auteurs avec des inventeurs du segment 1 et qu'ils aient hérité des citations reçues par ces-derniers.

**Tableau 7. Performance des brevets par segment**

	Nombre de brevets	Nombre de citations moyen par brevet
Segment 1	600	8.3
Segment 2	80	5.8
Segment 3	608	8.0
Segment 4	62	6.9
Segment 5	208	7.6
Segment 6	219	7.7

Après avoir analysé les brevets qui sont obtenus conjointement par des inventeurs du segment 1 et 3, nous observons effectivement que beaucoup d'inventeurs de ces deux segments sont des co-inventeurs. Sur les 608 brevets obtenus par les inventeurs du segment 3, plus de 341 (57%) ont été obtenu conjointement avec des inventeurs du segment 1. Il apparaît ainsi que les associations avec des inventeurs qui se font citer est une stratégie très avantageuse. Nous ne pouvons donc pas déduire que le coefficient de groupement est un important indicateur de performance en ce qui a trait au nombre de citations reçues. Il semblerait plutôt que faire parti d'un clique n'est avantageux qu'à partir du moment où au moins un des inventeurs du clique a un profil social axé sur la centralité et l'intermédiation et par ce fait produit des brevets de meilleurs qualités. Les inventeurs du segment 3 sont impliqués dans la production de brevets qui sont plus souvent cités parce qu'ils sont liés à des inventeurs qui ont une position centrale et d'intermédiaire plus important que les autres. Toutefois, lorsqu'ils sont associés avec des inventeurs du segment 2 (ce qui est le cas pour 37 brevets), ils se retrouvent impliqués dans la production de brevets de moins bonne qualité.

Finalement, nous observons que les segments 4, 5 et 6 ont des performances inférieures à ceux des inventeurs du segment 1. Cela est plutôt surprenant, puisque ceux-ci étaient très performants sur le plan du brevetage. Il semble donc que le degré d'intermédiation est un meilleur indicateur de performance que le degré de distribution lorsqu'il s'agit de prendre en compte la qualité des brevets produits.

## 5 CONCLUSION

Nous avons utilisé une technique de segmentation par densité pour regrouper les inventeurs suivant leur capital social. Cette méthode nous a permis d'identifier les inventeurs qui sont

périphériques, ceux qui font partie de cliques, ceux qui sont plutôt des intermédiaires et ceux qui ont des positions centrales au sein du réseau. Une analyse préliminaire de chacun des segments supporte l'hypothèse selon laquelle un capital social axé sur le degré de distribution et d'intermédiation a un impact positif sur la performance des inventeurs. Les inventeurs qui ont des profils sociaux de ce type ont un nombre de liens sociaux plus importants et évitent d'être confinés à des cliques. Pour ce qui est du coefficient de groupement et des cliques, cela semble être un trait social qui peut devenir avantageux à partir du moment où il implique une proximité avec des inventeurs centraux et intermédiaires. Ces résultats ont d'importantes implications au niveau de la gestion de l'innovation.

Pour les gestionnaires et stratèges corporatifs, il importe de prendre en compte les conclusions de la présente étude lorsque l'on sélectionne et développe un portefeuille technologique. Tout d'abord, il est préférable de sélectionner des projets dont au moins l'un des membres est un important technologue. Celui-ci devra avoir été impliqué dans de nombreux projets, avec différentes équipes ou départements et avoir eu un curriculum marqué par la diversité dans le parcours scientifique et technique. D'autre part, la stratégie de sélection et de financement devra favoriser le travail entre différents départements. Cette approche va permettre aux individus qui vont participer à ces projets de développer des liens sociaux diversifiés au sein de l'entreprise. Ceux-ci pourront alors devenir les membres productifs de futurs projets de développement.

Ces principes s'appliquent très bien à la réalité des grandes entreprises qui bénéficient d'un large éventail de ressources et de talent nécessaires à l'élaboration d'une stratégie technologique diversifiée. De plus, les grandes entreprises mettent souvent en place des départements d'intégrations. Nos observations suggèrent que ces départements jouent un rôle plus actif dans les projets technologiques importants, de manière à permettre à leurs membres d'élargir leurs liens sociaux au sein de l'entreprise.

Pour les décideurs au niveau public, il importe de mettre en place des politiques qui favorisent la diversification de l'activité industrielle, mais qui vont aussi encourager les collaborations interindustrielles. Au niveau de subventions octroyées pour le développement de nouvelles technologies, les projets qui impliquent de grandes entreprises ou qui impliquent une alliance entre plusieurs entreprises semblent être plus avantageuses. Dans le cas d'une PME, il est important de subventionner des projets qui sont menés par des technologues très actifs dans différentes communautés technologiques. Finalement, pour la subvention des universités, il est important de favoriser les projets qui sont développés en collaboration avec plusieurs départements.

De point de vue général, nous concluons le développement de nouvelles technologies doit avoir lieu dans des milieux multidisciplinaires parce que cela permet aux individus de bâtir des liens sociaux diverses et d'accumuler de vastes connaissances techniques. Que cela soit au niveau de l'entreprise, de l'industrie ou de toute une économie, disposer de ressources humaines qui ont un curriculum social et technologique diversifié est toujours bénéfique. Comme nous l'avons vu dans notre étude, les technologues qui ont développé ce type de profil innovent davantage et produisent des innovations plus importantes que les autres. Encourageons donc la diversité et la multidisciplinarité dans nos écoles, nos universités, nos entreprises et notre économie.

## 6 REMERCIEMENTS

Nous remercions le CRSH, l'IRSC et la CRSNG pour son support financier. Nous remercions l'OST et particulièrement Vincent Larivière pour l'accès et le support aux données concernant les citations de brevets USPTO. Nous remercions Ricard-Olivier Moreau et Terence-Shawn Ramine Kananian pour leur longue et laborieuse contribution dans le nettoyage et l'administration de notre base de données. Finalement, nous remercions Ilse-Carolina Moline Gutiérrez pour la préparation de la liste de mots clés en nanotechnologie.

## 7 RÉFÉRENCES

- Adler, P. S., Kwon, S.-W. (2002) Social capital - Prospects for a new concept. *The Academy of Management Review*, 27 (1), 17-40.
- Alencar, M. S. M., Porter, A. L., Antunes A. M. S. (2007) Nanopatenting patterns in relation to product life cycle. *Technological Forecasting & Social Change*, 74 (2007) 1661-1680.
- Amin, A., Cohendet, P. (2004) *Architecture of Knowledge: Firms, Capabilities and Communities*. NY: Oxford University Press.
- Ankerst, M., Breunig, M. M., Kriegel, H.-P., Sander, J. (1999) OPTICS: Ordering Points To Identify the Clustering Structure. *ACM SIGMOD international conference on Management of data*. 49-60.
- Berry, M. J., & Linoff, G. S. (2004). *Data Mining Techniques For Marketing Sales and Customer Relationship Management*. Wiley.
- Burt, R. S. (2004) Structural Holes and Good Ideas. *The American Journal of Sociology*, 110 (2), 349-399.
- Cantner, U., Graf, H. (2006) The network of innovators in Jena: An application of social network analysis. *Research Policy*, 35, 463-480.
- CodePlex (2001) <http://nodex1.codeplex.com/>.
- Cohen, W., Levinthal, D. (1990) Absorptive capacity: a new perspective on learning and innovation. *Administration Science Quarterly*, 35, 128-152.
- Cowan, R., Jonard, N. (2004) Network structure and the diffusion of knowledge. *Journal of Economic Dynamics & Control*, 28, 1557-1575.
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X. (1996) "A density-based algorithm for discovering clusters in large spatial databases with noise". Simoudis, E., Han, J., Fayyad, U. M. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 226-231.
- Fitzgibbons, K., McNiven, C. (2006) Towards a Nanotechnology Statistical Framework. *Blue Sky Indicators Conference II*.
- Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78 (6), 1360-1380.
- Griliches, Z. (1990) Patent Statistics as Economic Indicators - A Survey. *Journal of Economic Literature*, 28 (4), 1661-1707.
- Harhoff, D., Narin, F., Scherer, F. M., Vopel, K. (1999) Citation Frequency and the Value of Patented Inventions. *The Review of Economics and Statistics*, 81 (3), 511-515.
- Inkpen, A. C., Tsang, E. W. K. (2005) Social capital networks, and knowledge transfer. *Academy of Management Review*, 30 (1), 146-165.
- Manning, C. D., Raghavan, P., Schütze, H. (2008) *An Introduction to Information Retrieval*, Cambridge University Press.
- Mogoutov, A., Kahane, B. (2007) Data search strategy for science and technology emergence: A scalable and evolutionary query for nanotechnology tracking. *Research Policy*, 36, 893-903.
- Nahapiet, J., Ghoshal, S. (1998) Social Capital, Intellectual Capital, and the Organizational Advantage. *The Academy of Management Review*, 23 (2), 242-266.
- Newman, M. (2003). The structure and function of complex networks. *SIAM Review*, 45 (2), 167-256.
- Porter, A. L., Youtie, J., Shapira, P., & Schoeneck, D. J. (2008). Refining search terms for nanotechnology. *Journal of Nanoparticle Research*, 10, 715-728.
- Rapid-I (2011) <http://rapid-i.com/>.
- Schmoch, U., Heinze, T., Hinze, S., Rangnow, R. (2003) *Mapping Excellence in Science and Technology across Europe: Nanoscience and Nanotechnology*. Centre for Science and Technology Studies.
- USPTO (2009) <http://uspto.gov>.
- Wasserman, F., Faust, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- Watts, D., Strogatz, S. (1998) Collective Dynamics of 'small-world' networks. *Nature*, 393, 4 June, 440-442.
- Zitt, M., Bassecoulard, E. (2006) Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information Processing and Management*, 42, 1513-1531.
- Zucker, L., Darby, M., Armstrong, J. (1998) Geographically localized knowledge: Spillovers or markets? *Economic Inquiry*, 36(1), 65-86.