Titre: Title:	Collaboration spaces in Canadian biotechnology: A search for gatekeepers
Auteurs: Authors:	Andrea Schiffauerova et Catherine Beaudry
Date:	2012
Référence: Citation:	Schiffauerova, Andrea et Beaudry, Catherine (2012). Collaboration spaces in Canadian biotechnology: A search for gatekeepers. *Journal of Engineering and Technology Management*, *29*(2), p. 281-306. doi:10.1016/j.jengtecman.2012.03.004

Titre de la revue: Journal Title:	Journal of Engineering and Technology Management
Maison d'édition: Publisher:	Elsevier
URL officiel: Official URL:	http://dx.doi.org/10.1016/j.jengtecman.2012.03.004
Mention légale: Legal notice:	In all cases accepted manuscripts should link to the formal publication via its DOI

Gatekeepers of Canadian biotechnology clusters – Where geographical clusters and co-invention networks intersect[±]

Andrea Schiffauerova

Department of mathematics and industrial engineering, École Polytechnique de Montréal, Canada

Catherine Beaudry[±]

Department of mathematics and industrial engineering, École Polytechnique de Montréal, Canada
Center for Interuniversity Research on Science and Technology (CIRST)

[±] Beaudry acknowledges financial support of the Social Science and Humanities Research Council of Canada. We acknowledge helpful comments from Jorge Niosi and Nathalie de Marcellis-Warin, from the InnovaRisQ research team as well as from four anonymous referees. We are grateful to Ahmad Barirani for the extraction program used in the data collection. None of these, however, are responsible for any remaining errors.

[±] Corresponding author: Department of mathematics and industrial engineering, École Polytechnique de Montréal, C.P. 6079, succ. Centre-ville, Montréal (Québec) H3C 3A7, Canada, tel. : +1 514 340-4711 ext 3357, fax : +1 514 340-4173, catherine.beaudry@polymtl.ca

Gatekeepers of Canadian biotechnology clusters – Where geographical clusters and co-invention networks intersect

Abstract: The paper studies the network of Canadian biotechnology co-inventors and compares the structure of the links between inventors within, and outside of, the cluster. Two types of proximity between individuals are compared: within cluster co-invention refers to geographical proximity while co-patenting links (network component) represent social proximity between inventors of an epistemic community. We show that the cluster-based subnetworks are more fragmented and less centralized than the network components. The paper then proposes two indicators measuring an inventor's importance as a gatekeeper, *i.e.* the individuals at the cluster's frontier responsible for the inflow of the external knowledge to the cluster.

Key words: gatekeepers, innovation network, knowledge transmission, geographical cluster, biotechnology

JEL Classification: O31, R12, D85

1. Introduction

There has been a recent debate about the fact that geographic proximity might not be a universal panacea for the transmission, adoption and generation of knowledge (Boschma, 2005; Wink, 2008). Although geographical proximity facilitates knowledge sharing and spillovers are generally geographically bound (Audretsch and Feldman, 1996; Jaffe *et al.*, 1993), a number of other types of proximities can substitute geographical proximity. A number of scholars for instance argue that it is cognitive proximity[1] that causes tacit knowledge to spill over between firms (Cowan *et al.*; 2000; Breschi and Lissoni, 2001a, 2001b) whether within a close distance or over a geographically dispersed epistemic community, *i.e.* that shares a common science base. As Breschi and Lissoni (2001a, p. 989) suggest, “it is physical [or geographical] proximity that follows epistemic proximity, and not vice versa”. Building from the work of the French Proximity Dynamics group (see for instance Torre and Gilly, 2000) on space and proximity, Boschma (2005) suggests that their definition of organizational proximity encompasses a combination of cognitive proximity and social proximity. As a consequence, just sharing the same knowledge base may not be enough for knowledge transmission to take place, social proximity is also required.

Indeed, social proximity combined with geographic proximity results in more effective knowledge transfer (Owen-Smith and Powell, 2004; Sorenson and Stuart, 2001). Members of these socially proximate communities can be divided into different categories depending on the types of links they maintain within a network. A great deal of attention has been devoted to intermediaries that bring together groups of members that would have no interactions otherwise

---

[1] Epistemic proximity for Breschi and Lissoni (2001a, 2001b) is somewhat similar to what Boschma (2005) refers to as cognitive proximity, a concept developed by Nooteboom (1999).

(Hargadon and Sutton, 1997; Marsden, 1982). In fact, it is suggested that the intermediaries or brokers that bridge over structural holes (Burt, 1992, 2007) gain efficiency, have a greater influence (Fernandez and Gould, 1994) or make more profitable alliances (Stuart *et al.*, 2007). Gould and Fernandez (1989) name *gatekeepers* the brokers that transmit knowledge received from outside to the group to which they belong.

There is however a lack of research on individuals as gatekeepers carried out in a more global context as most studies focus either on the role of the gatekeeper-firms within a supply chain or a geographical cluster context or on the role of gatekeeper-individuals within a company, but the whole national network of these individuals together with all their intra-cluster and inter-cluster connections has not been taken into consideration. We therefore pose two main research questions in this paper: Who are the key individuals that enable the potential nurturing of geographical clusters[2] with fresh external knowledge? How can these gatekeepers be identified in the national network of inventors and how can their importance for the geographical cluster and for the country be evaluated? This paper intends to provide answers to these questions.

Building on this theoretical framework, we aim to develop a method to systematically identify the individuals that are at the frontier of geographical clusters and that while being connected within the cluster (geographic proximity), also have collaborative links (social proximity) outside its boundary within an epistemic community. The network of Canadian biotechnology innovation is built from the co-patenting links between inventors of an epistemic

---

[2] In this study, a geographical cluster is defined as a geographically continuous region active in biotechnology (as measured by the patent production).In addition, clusters will be assumed to be geographical clusters. In order to lighten the text, every time the word cluster is used, a geographical cluster will be implied.

community. The address of each inventor allows us to locate them with clusters and thus to establish the geographical proximity between these individuals.

This paper compares and discusses the structure of the cluster-bound[3] co-invention subnetworks with that of the component-based[4] co-invention subnetworks, and investigates the level and nature of the overlap between them. Figure 1 illustrates the interaction between the cluster-based subnetwork (inventors A to E, J and L) and the component-based subnetwork (inventors A to H). The individuals responsible for the interaction between the two collaboration subnetworks are identified and their importance for the cluster highlighted. These are our gatekeepers - the inventors who connect out-of-cluster inventors with within-cluster inventors and thus potentially enable the nurturing of biotechnology clusters with fresh knowledge originating outside. We find that only around 10-20% of inventors from each cluster can be categorized as gatekeepers but are potentially critical for the inflow of external knowledge to the cluster. This paper presents a systematic way to identify them and to determine their relative importance as potential procurers of external knowledge for clusters and for Canada. Because they span over structural holes (Burt, 1992), these are the individuals that biotechnology firms should be recruiting as an essential part of larger teams of researchers. Similarly, if universities are to successfully commercialize some of their innovation ideas, the academics behind the research should also be well connected if not gatekeepers themselves. The methodology

---

[3] A cluster-bound subnetwork only considers the links between inventors that are in close geographic proximity, i.e. within a cluster.

[4] A network is composed of distinct components, within which the nodes are linked directly or indirectly, which have no links between one another.

described in the paper provides a relatively simple and effective way to identify the most important knowledge gatekeepers.

(Insert Figure 1 ***here)***

The paper is organized as follows: section 2 reviews the relevant literature on brokers, gatekeepers, geographic and other types of proximities, section 3 introduces the data and methodology, section 4 presents the results related to collaboration in the geographically-bound co-invention subnetworks with that of the component-based co-invention subnetworks including a description of the main network structure properties used in this study, section 5 explores the points of interaction between them, section 6 discusses the results, section 7 describes the limitations of this study and how we plan to overcome these limitations in the future and finally section 8 concludes.

2. Theoretical background

Over the last two decades there has been an emerging interest in the role of intermediaries (brokers) in the innovation process. Brokers are either individuals or organizations who "facilitate transactions between other actors lacking access to or trust in one another" (Marsden, 1982, p. 202). By enabling the flow of resources between otherwise unconnected groups, the brokers assume an important role in innovation networks and thus received plenty of attention from the research community.

One of the most widely acknowledged works on brokers is Burt's (1992) theory of structural holes, which describes how firms embedded in sparse networks of disconnected partners gain efficiency and control benefits. A "structural hole" is a gap in the flow of information between subgroups in a larger network. A firm occupying many structural holes has an advantage over competitors, because it has an easier access to information (due to many non-redundant contacts) and a greater control over the flow of information between disconnected partners. The empirical research has confirmed the power that the brokers gain due to their network positions. Fernandez and Gould (1994) show that organizations which occupy brokerage positions in the national health policy domain are more likely to have greater perceived influence. Stuart *et al.* (2007) show that biotechnology firms acting as brokers have higher chances to make profitable alliances with downstream partners.

Brokers may also be individuals and a number of studies examine their influence within particular organizations. Burt (2004) finds that individuals who span structural holes in an organization gain substantial social capital (compensation, positive performance evaluations, promotions, etc.). In a later article, Burt (2007) nevertheless points out that the brokerage benefits are dramatically concentrated in the immediate network around a broker, but that the benefits are much reduced in case of second-hand brokerage (transfer of information between people with whom a broker has only an indirect connection). Hargadon and Sutton (1997) study technology brokering of engineers, designers and managers of a product design firm, IDEO, and summarize the brokering process as knowledge access, acquisition, storage and retrieval. One could argue that IDEO is in a certain way a broker because of the brokering role taken by its designers.

Winch and Courtney (2007) describe the role of organizations specifically founded to undertake an intermediary role, or innovation brokers, as the transfer of knowledge between the sources of new ideas and the users of those ideas in innovation networks. Hargadon and Sutton (1997) however stress that brokering is more than just transferring knowledge, a broker also serves as a repository of knowledge, which allows him (the individual within the firm and, by extension in their case, the firm as well) to recombine existing ideas from various resources and to generate solutions to the problems in other industries. As a consequence, knowledge brokers in nanotechnology, for example, have to develop high absorptive and transformative capacity to be able to disperse the knowledge generated by the innovation system (Pandza and Holt, 2007).

Obviously, the brokerage role is quite varied, and brokers can facilitate transactions in a number of distinct ways (see Howells, 2006[5] for a description of the functions undertaken by intermediaries). In their classification of brokerage roles, Gould and Fernandez (1989) identify five types of brokers based on the network configurations that result when a broker connects two otherwise unassociated partners: coordinators, itinerant brokers, gatekeepers, representatives and liaison officers. For instance, the partners and broker can come from the same organization in which case, the broker is a coordinator, or from three distinct organizations, in which case, the broker is considered a liaison officer between the knowledge transmitter and the knowledge receiver. Among these five possible roles, a broker acts as a gatekeeper when belonging to the receiving group but not to the transmitting group.

However, it is not only private firms that assume gatekeeping functions, but also research universities and cooperative R&D institutions (Steiner and Ploder, 2007). Public research organizations have been even suggested to serve the functions of a gatekeeper to a higher degree than private actors (Graf, 2008).

Wink (2008) considers individuals as well as firms and institutes as gatekeepers during the knowledge generation process. Knowledge generation and examination (the first two steps of the knowledge generation process) require gatekeeping activities from individual researchers while research based firms and institutes concentrate on knowledge exploitation. To our knowledge, this study is the closest research on gatekeepers that has attempted to follow a multilevel

---

[5] For Howells (2006), brokering and gatekeeping are only one category of the typology of intermediary roles in the innovation process. In this paper, we will consider gatekeeping in a more narrow sense.

approach in the sense intended by Klein and Kozlowski (2000). Wink (2008) is an exception as scholars generally examine the role of firms as intermediaries, but relatively few examine the individuals who perform this intermediary role within the context of a particular organization. It was Allen (1967) who first identified certain industrial researchers in an organization as key persons in the innovation process, because they gather, process and transfer information from internal and external processes. These individuals were labelled gatekeepers. It has been shown (Allen, 1977; Tushman and Katz, 1980; Katz and Tushman, 1981) that the total performance of the R&D system in the firm is in fact critically dependent on a few key individuals, the gatekeepers, because they provide a linking mechanism between the company and its external environment.

The role of gatekeeper has also been studied at the geographical cluster level of analysis, i.e. between local firms and non-local firms. In that case, gatekeepers are generally characterized as leading firms that search for non-local knowledge, transmit it into the region and thus link the region with the outside world (Morrison, 2008). In Gould and Fernandez's (1989) typology, the characteristic that would group the gatekeeper and the knowledge receiver would be that they are both located within the same region. Leading firms can act as gatekeepers not only due to the well-established external contacts, but also due to their superior knowledge base, technological resources and capabilities that make them better equipped to absorb new knowledge and facilitate its diffusion throughout the geographical cluster (Malipiero *et al.*, 2005). The absorptive capacity of the gatekeepers is also at the heart of the research of Lazaric *et al.* (2008) who propose the way of its effective realization, while Boschma *et al.* (2007) study the impact of the local network positions of the firms and their connectivity to the non-local firms on their innovative performance. This highlights the importance of the point of connection between geographically proximate collaboration and distant co-inventorship.

Economic geography often views regions as key drivers of innovation. This is built on the suggestion that *geographical proximity* facilitates knowledge sharing, since knowledge does not spill over large distances (Audretsch and Feldman, 1996; Jaffe and Trajtenberg, 1996). It is assumed that all firms in the cluster can benefit from these localized knowledge spillovers, which are not available to the firms outside the clusters. As a consequence, the firms in geographical clusters are found to be more innovative (Baptista and Swann, 1998; Beaudry, 2001, Beaudry and Breschi, 2003). However, Waguespack and Birnir (2005) show that patents co-invented by inventors located across more than one state in the US receive the highest rate of citations hence suggesting faster knowledge diffusion. As a consequence, Boschma (2005) suggests that this view overemphasizes the role of geographical proximity in the transfer of knowledge between firms. He argues that other dimensions of proximity should be taken into consideration as well, since geographical proximity per se is neither a necessary nor a sufficient condition for organizational learning to take place. Other types of proximities can act as substitutes of geographical proximity and yield benefits that may have falsely be solely attributed to, and facilitated by, geographical proximity.

Wink (2008) proposes that gatekeepers can provide interface nodes between regional innovation systems by different forms of proximity. The ability of an actor to function as a gatekeeper thus depends on the kind of proximity which is necessary to span the boundary between the systems. There are several dimensions of proximity described in the literature. Torre and Gilly (2000) make a distinction between two different dimensions: geographic proximity, which refers to the spatial context, and organizational proximity, which is based on the organizational interaction of firms participating in clusters (and includes a cognitive dimension as well). Kirat and Lung (1999) incorporate a third dimension, institutional proximity, indicating that the closeness among the agents is influenced and restricted by the institutional environment.

Boschma (2005) extends this classification and identifies five dimensions of proximity – cognitive (proximity related to the knowledge base of the actors), organizational (closeness of actors in organizational terms), social (closeness based on the socially embedded relations between agents, which involve trust, friendship, kinship and experience), institutional (proximity related to the institutional environment) and geographical (defined as the spatial or physical distance between economic actors).

Literature on knowledge creation and diffusion emphasizes the role of cognitive proximity. It is argued (Cowan *et al.,* 2000; Breschi and Lissoni, 2001a, 2001b; Lissoni, 2001) that it is not geographic proximity that causes tacit knowledge to spill over between firms, but it is social connectedness of members of the same epistemic community in the network, *i.e.* the social proximity of individuals that are close cognitively. Breschi and Lissoni (2001a) further suggest that members of epistemic communities develop a common language or codebook that prevent local actors who live and work close to community members from understanding the messages exchanged within the epistemic community, whether locally or over large distances. Knowledge circulates and flows through the network between actors who are not necessarily placed in the same location. Technical or scientific knowledge is highly specific and its jargon differs from the jargon of the broader social community. Those who understand it are the members of closed, restricted, but geographically dispersed epistemic community, within which tacit messages can be easily transmitted even if knowledge links take place among agents located far away in space. The networks thus do not require co-location of the actors for the production of innovation. Physical proximity does not however imply epistemic proximity, because epistemic communities are never as wide as to include all members of a local community. This means that firms in clusters may be excluded from knowledge sharing when they are not part of knowledge networks.

Apparently, the two concepts seem to stand against each other. Does it matter more for an inventor to be in the right location or to be connected to the right network of people? It has been argued that the combined effects of geographic and social spaces result in a more effective knowledge transfer (Owen-Smith and Powell, 2004; Sorenson and Stuart, 2001), while the causal relationship between the geographical and social distances has been suggested as well (Sorenson, 2003). We believe that both the concept of space and the concept of network are of the utmost importance for the knowledge creation and diffusion. Both geographical and social dimensions (and to a certain extent the cognitive dimension as well) potentially nurture the growth of the geographical cluster and promote innovation through a dynamic interaction of actors localized in these clusters who absorb external knowledge through local and non-local networks. In order to bring new knowledge to the geographical cluster, gatekeepers thus have to be well connected both inside and outside these clusters.

From the analysis of the studies mentioned above, we have noted a lack of research on individuals as gatekeepers carried out in a more global context. In the introduction, we raised two questions that can be summarized as: Who are the gatekeepers at the frontier of geographical clusters? How can they be identified and their importance measured? To answer these two questions, this paper will build the co-patenting network of Canadian biotechnology inventors and examine the importance of individual gatekeeper-inventors in knowledge transmission between two subnetworks – the cluster-based subnetwork and the part of the network component that is out of the cluster. The *cluster-based subnetwork* in this context is based on geographic proximity and characterized by co-location of biotechnology inventors in the 12 most important Canadian cities in terms of biotechnology patent production (the geographical clusters). It assumes that co-inventorship links are mostly geographically localized and that no significant out-of-cluster linkages exist. Indeed, it has been shown that majority of all collaborative activities in Canadian biotechnology are carried out within clusters (Schiffauerova and Beaudry, 2009). The *network component* consists of all the inventors that are directly and indirectly connected. A network is therefore composed of a number of distinct components that do not have any links between one another. These components can therefore represent the epistemic communities mentioned in the literature. A component is then composed of the individuals (and their links) that are co-located in the same cluster as well as their out-of-cluster co-inventors. The individuals that have connections both within the cluster and outside the cluster are thus potential gatekeepers. From its construction, a network component necessarily accounts for social proximity (because the co-inventors of a patent have worked together and hence know each other). There must therefore be a minimum level of trust and experience involved in the collaboration towards the development of a patent. It also somewhat relates to cognitive proximity as the research teams that co-invent these biotechnology products and processes partly share a common knowledge base. The paper will however focus only on two proximity dimensions relevant to the identification of individual gatekeepers: geographical and social[6].

---

[6] Following the French Proximity Dynamics group mentioned in introduction, we could say that our research focuses on geographical and organisation proximities, where according to Boschma (2005), the latter comprises both social and cognitive proximities. Our research does not attempt to measure cognitive proximity as such. We assume that all inventors who have collaborated on biotechnology innovations with each other at some point and are thus directly or indirectly interconnected in a network component are also part of the same epistemic community*, i.e.* restricted by the scientific fields and technological specializations. Measuring cognitive distances between components or individuals would require to identify biotechnology subfields and their technological closeness which goes beyond the scope of this research. We thus prefer to use Boschma's taxonomy of proximities and focus on the two types we can accurately measure.

3. Methodology

3.1 Data

The data used for the empirical analysis is the United States Patents and Trademarks Office (USPTO) database. The choice of the USPTO database rather than the Canadian Intellectual Property Office (CIPO) stems from the fact that the latter does not contain the geographical location of each inventor. The use of the USPTO database instead of the CIPO may introduce a certain bias in the data. We consider this bias minimal, as Canadian inventors usually patent both in Canada and in the United States since the much larger and easily accessible American biotechnology market offers a greater potential than that of Canada.

Biotechnology encompasses several different research technologies and several fields of application. We have thus opted to ground our USPTO search strategy on the OECD definition of biotechnology, which is based on a group of carefully selected International Patent Classification (IPC)[7] Codes. An automated extraction program was used to collect the required information such as patent numbers and inventors' names and addresses from biotechnology patents. All biotechnology patents granted before March 31, 2007 were included (the first of such patent being granted in 1976). Within these IPC codes, there are around 100 000 biotechnology patents registered at the USPTO. We created a patent database, which contains all patents for which at least one inventor resides in Canada, and which comprises 3550 patents. Following the concept of social networks, we created connections between the inventors from the extracted patent information and constructed the Canadian biotechnology innovation network using the social

---

[7] The OECD definition of biotechnology patents covers the following IPC classes: A01H1/00, A01H4/00, A61K38/00, A61K39/00, A61K48/00, C02F3/34, C07G(11/00, 13/00, 15/00), C07K(4/00, 14/00, 16/00, 17/00, 19/00), C12M, C12N, C12P, C12Q, C12S, G01N27/327, G01N33/(53*, 54*, 55*, 57*, 68, 74, 76, 78, 88, 92).

network analysis program PAJEK. Our social network comprises of inventors (vertices or nodes) that are linked by their co-invention relationships with other inventors (edges or lines), i.e. when they have co-invented an innovation that leads to a patent. The analysis of this network enables us to describe its structural properties and to explore the collaborative behavior of inventors inside and outside Canadian biotechnology clusters.

Since the patent data providing the connections between inventors span over a period of 31 years, we assume that once inventors collaborate on one patent, they continue to be in contact afterwards and are able to exchange information with all their collaborators long after the patent has been granted. Common wisdom suggests that it takes about ten years for a human health biotechnology product to be on the market from its inception. Hence, inventors are probably involved in the project for a large proportion of this period. Dahl and Pedersen (2005, p. 89) suggest that "the relationships created through formal projects persist even after the project. Project participants remain in social contact, which increases the probability that knowledge is shared." Because 99% of the network components are composed of a relatively small number of individuals and exist for a small number of years, we can safely disregard the time of collaboration and consider all links among inventors in the network as active 'simultaneously'. We are conscious that this may appear as a strong assumption and are aware of the limitations that it may entail. It is however important to note that the network is composed of a large number of disjoint network components that do not span the entire 31 years of the database but much shorter periods of time. For the purpose of our analysis, the time dimension is therefore not crucial. One has to keep in mind that the gatekeepers identified are not all 'active' at the same time. This paper does not aim to examine the dynamics of gatekeeping activities but to identify the importance of the gatekeepers. This is however a path that we intend to pursue in the future.

3.2 Network structure properties

Throughout the paper we will be using several measures of the network structure properties, whose basic grasp is necessary for understanding the discussed concepts. Their brief description, based on Wasserman and Faust (1994) and de Nooy *et al.* (2005) follows.

Suppose a very simple network composed of five inventors (vertices) and their collaborative links (Figure 2). *Structural cohesion* within a network refers to the degree to which vertices (inventors in our case) are connected among themselves. Usually it is measured by the *density of a network* (which is the number of existing lines in the network expressed as a proportion of the maximum possible lines). In our example, there are six lines out of a potential of ten connections. To compare networks of very different sizes, the *average degree of a network* (degree of a vertex is the number of lines that are incident with it, i.e. that are directly connected to the vertex) is generally used because it is not affected by network size. In our example, A, B and C are connected to three other inventors, E is connected to two, and D to only one other. The average degree would therefore be 2.4.

(Insert Figure 2 here)

A shortest path between two vertices is referred to as geodesic. The *geodesic distance* is then the length of a geodesic between them, which depends on the number of steps (or links) needed for an inventor to reach another inventor in the subnetwork. In our example, the geodesic between A and D is 2 (and the path goes via C, from A to C and from C to D). A short path length in innovation networks should improve knowledge production and knowledge diffusion (Cowan and Jonard, 2004; Fleming *et al.*, 2004), since knowledge can move to the different parts of a network more quickly and spread rapidly among inventors. The longest geodesic in a network (the longest shortest path) is called *diameter of a network*. In our example, the longest shortest path is 3 connecting E and D (via B and C, from E to B, from B to C and from C to D, or via A and C, from B to A, from A to C and from C to D). A more global measure is *average distance of a network* (measured only in a connected network) which takes the average of all geodesic distances. The *reach of a vertex* is defined as the number of vertices that can be reached from the vertex, both directly and indirectly.

The *centrality of a vertex* indicates whether the position of an individual inventor within the subnetwork is more central or more peripheral. Inventors that are more central have better access to information and better opportunities to spread information. We measure both *degree centrality* (which equals the degree of a vertex defined above) and *betweenness centrality* (a proportion of all shortest distances between pairs of other vertices that include this vertex). The latter indicates the importance of a vertex as an intermediary in the network. Betweenness centrality involves counting the number shortest distances between all other vertices but A for instance. Between B and C, the shortest distance does not involve A, because they are directly connected. The same can be said for B and E as well as for C and D. The shortest distance between D and E as well as between C and E goes through either B or A. Only two of the shortest distances would involve A.

Centralization characterizes an entire network. A highly centralized network has a clear boundary between the center and the periphery. The center of the centralized network allows more efficient transmission of information, which consequently spreads fairly easily in highly centralized networks. We use two measures of the network centralization, all based on the variation in centrality of all vertices in a network: *degree centralization* and *betweenness centralization*. The former will be higher when a network has a clear center through which most "traffic" goes, so to speak. The latter measures the heterogeneity of the network in terms of the importance of intermediaries. If all vertices act as intermediaries, the variation will be small and so will the betweenness centralization.

Cliquishness[8] is a property of local network structure which refers to the likelihood that two vertices that are connected to a specific third vertex are also connected to one another. Cliquish networks have tendency towards dense local neighborhoods, in which individual inventors are better interconnected with each other. Such networks exhibit a high transmission capacity, since a great amount of information could be diffused rapidly (Burt, 2001). In this paper we measure the degree of local cliquishness for each vertex with the egocentric density of a vertex (which is the fraction of all pairs of the immediate neighbors of a vertex that are also directly connected to each other). In our example, A has three immediate neighbors, B, C and E. The egocentric density of A thus refers to the fact that B and C are connected and so are B and E, but C and E are not. The cliquishness of a network is then calculated by the *average egocentric density of the network* over all vertices.

As mentioned above, a network is composed of a number of disjoint *components* of vertices that are linked directly or indirectly, but the components have no connections among one another. The main measures regarding components include the size of the largest components in a network, *i.e.* the number of inventors (vertices), the average component size in a network and the number of isolated vertices (1-inventor components).

---

[8] Cliquishness is also referred to in the literature as "clustering" but we will not use this terminology so as to not confuse the reader with the geographical clusters.

4. Collaboration

4.1 Collaboration in the geographical space

The network of Canadian biotechnology inventors includes 4569 inventors (vertices) and 9731 collaborative relations[9] (edges). Based on the location of inventors we have identified 12 Canadian biotechnology clusters: 20% of inventors reside in the Toronto cluster, 15% in the Montreal cluster and 9% in the Vancouver cluster. Only a very small portion of Canadian inventors live outside the defined clusters (around 3%) and around 29% of inventors in our sample reside outside the Canadian borders.

Knowledge spillovers[10], a supply-side benefit, are often discussed in the context of biotechnology innovation. The fact that biotechnology knowledge is largely tacit limits knowledge diffusion over long distances. As the transmission of tacit information and knowledge spillovers is usually associated with face-to-face contact, the collaboration among inventors working in geographical clusters is encouraged by the benefits of acquiring the knowledge which the subjects located in close geographical proximity spill over. This section of the paper analyzes local collaborations carried out entirely within geographical clusters, and as such we divide the Canadian biotechnology innovation network into geographically bound cluster subnetworks, hence ignoring or "severing" all relations outside of the cluster. Each subnetwork strictly includes inventors located in that particular cluster, while excluding the ones that are not. We then study

---

[9] Each collaborative relation (also called a collaborative link) represents a connection between a pair of inventors, which involves one or more instances of co-invention of a biotechnology patent.

[10] Following Breschi and Lissoni (2001b), we define localized knowledge spillovers as knowledge externalities bounded in space that allow companies operating nearby key knowledge sources to introduce innovations at a faster rate than rival firms located elsewhere.

the structure of these subnetworks and postulate on how knowledge is potentially transferred throughout.

Table 1 presents some of the main structural properties of the subnetworks created in this manner. In previous work, we have examined the network architectures of each cluster and related them to the efficiency of each subnetwork in terms of knowledge diffusion and innovation creation (Schiffauerova and Beaudry, 2009). We propose that for the network to be efficient at knowledge transmission and generation it should be cohesive (which means that inventors are closely interconnected), cliquish (which fosters trust and close collaboration), have a long reach within large components (which should facilitate the input of new and non-redundant knowledge from distant locations) and have a centralized structure (which supports fast information transmission). We find however that the structural subnetwork properties within each individual geographical cluster are quite diverse throughout Canada.

(Insert Table 1 here)

The cluster-based subnetworks are rather fragmented. Even though collaboration within clusters generally involves a very short geographical distance (commuting distance), inventors often choose to work in isolated groups. The fact that the largest components contain only 9%-18% of all inventors in each geographical cluster confirms that inventors co-located within the same cluster are not highly interconnected. Furthermore, a substantial part of the collaborative links is directed outside the cluster. Beaudry and Schiffauerova (2009) find that Canadian inventors frequently take part in joint research projects including collaborators from abroad (29% of collaborations[11]) or their colleagues located in other clusters (11% of collaborations). It would

---

[11] Collaboration here means a connection between a pair of inventors for the purpose of co-invention of one biotechnology patent. Each collaborative link may thus involve one or more collaborations.

appear that the social proximities are quite important for biotechnology invention in Canada. The following section therefore disregards the geographical proximity and focuses solely on the technological network.

4.2 Collaboration in the network

Within the network, collaboration is based on network components. As explained above, all inventors in a component are directly or indirectly interconnected and it is thus supposed that they all collectively contribute to the innovation process. The attachment of inventors to their local environment is considered as secondary and the innovation network is analyzed regardless of the inventors' location.

Canadian biotechnology inventors are grouped into 894 components, which suggests that the network is quite fragmented and that inventors are not highly interconnected (see the second column of Table 1 in the top part). Of the 894 components, the 30 largest are presented in Table 2. In terms of the number of vertices, the largest component (Component C1) includes 579 inventors, the second one (Component C2) consists of 185 inventors and the third (Component C3), of 175 inventors. There are few large components (10% of components include around 50% of inventors); most components however are relatively small. As a consequence, the average number of inventors in a component is also relatively small (5.11). This is attributable to the fact that around 22% of all components (195 components) are isolates (components that consist of sole inventors who have not collaborated), which represents 4% of inventors.

It is obvious that most components consist of inventors residing in several distinct clusters (in Table 2). This is particularly true for the largest components, where inventors are geographically spread over the entire country and abroad (Components C1 or C2). Some components, however, clearly consist of a great majority of inventors located in one cluster. For instance, Component C3 seems to incorporate inventors from five Canadian clusters, but a closer inspection shows that 112 out of 124 Canadian inventors of Component C3 come from Montreal. Because the largest Montreal's component has 109 inventors[12] (see Table 1), only 3 Montrealers are disconnected from the component if no inventors from other clusters or regions were included. These out-of-cluster inventors are the individuals that link the three lone Montreal inventors to the rest of Montreal-based C3 inventors. Similarly, 75 inventors of the largest component of Ottawa collaborate in Component C2, which includes a total of 77 Ottawa inventors. Two inventors are therefore connected to their fellow Ottawa residents indirectly via outsiders. Most of the other clusters' largest components are contained within Component C1, which looks like a great collaboration field for the most connected Canadian researchers except those from Montreal and Ottawa. In the case of Ottawa, we suspect that this is caused by the federal research concentration of the National Research Council seated in the Canadian capital, but Montreal is quite surprisingly isolated from the largest Canadian collaboration group of Component C1.

(Insert Table 2 here)

Some components (C6, C7, C19, C23 or C28) present intra-cluster collaboration, but also include some foreign collaboration relations. In fact, all of these 30 largest components include at least one foreign collaborator. Some of these mainly foreign[13] components consist of a majority

---

[12] The maximum reach for the Montreal subnetwork is 108 inventors, implying that each inventor in the largest component of that cluster subnetwork can reach 108 other inventors, hence the size of the largest component is 109 (108 + 1) inventors.

[13] An obvious limitation to this study is the degree to which a component is international or foreign. Because we have only extracted the patents to which at least one Canadian inventor has contributed, we omit the international part of the world biotechnology network that does not involve Canadian inventors.

of foreign inventors with only one or two Canadians (Components C10 or C14). These are probably much larger foreign networks to which a few Canadian inventors participate. For instance, Component C10 is based on collaboration on one single patent and is composed of 24 inventors; out of which 23 are foreign and only one is Canadian. Understandably, these mostly foreign components also show very low ratios of patents per inventor.[14]

Let us now turn to the network characteristics of these 30 largest network components (Table 3). Four largest components usually show higher cohesion and lower centralization than smaller components. They obviously also have larger geodesic distances but higher maximal reach, since it takes longer for the information to travel all over the large component but it can reach many more other inventors. Striking exceptions to this pattern are two medium-sized components, in which all inventors (Component C14) or almost all inventors (Component C10) are connected to each other, since they have all collaborated with each other on all their patents (or almost all for Component C10). The larger components may however consist of several smaller components connected by a few individuals.

(Insert Table 3 here)

A comparison of the structural properties with the cluster-based subnetworks (Table 1) reveals that the component-based subnetworks (or simply the components) (Table 3) are denser, more centralized and present more cliquishness, but they also have greater diameters. This should not be surprising as the cluster-based subnetworks are in facts smaller parts of components isolated by cluster boundaries. Hence collaboration within components is probably more efficient

---

[14] We are well aware of the fact that concentrating on inventors of Canadian patents may miss some much larger North American or even worldwide network which might link (indirectly) some of the components obtained. Since our focus is on Canadian cluster gatekeepers, however, this does not constitute an obstacle to our study.

because higher structural cohesion of subnetworks indicates closer interconnectedness of inventors, higher cliquishness fosters trust and close collaboration, and higher centralization supports fast knowledge transmission. In contrast, the cluster-based subnetworks show smaller diameters due to the high structural fragmentation. This means that the paths are shorter and information can travel faster in cluster-based subnetworks, but because of the smaller maximal reach, knowledge could potentially be acquired by much less inventors.

It is not unexpected that the transmission of knowledge through the network is more efficient if there are no geographical barriers and all the interconnected inventors could freely and frequently cooperate regardless of the distance between them. In reality, however, this is not usually the case. Even though we observe that collaboration of Canadian inventors with non-local partners are very common in biotechnology, for most inventors, in fact, local intra-cluster collaborative relations are more frequent (Beaudry and Schiffauerova, 2009). Biotechnology inventors in Canada do take the geographical distance into consideration when searching for partners. Consequently we consider both the technological network and the geographical clusters as extremely important concepts and our final task is thus to seek the points of interaction between the two. Since our cluster-based subnetworks consist of the local fragments (within the geographical cluster) of the component-based subnetworks, let us now find the key individuals who link the former to their out-of-cluster co-inventors.

5. In a search of the gatekeepers

This last part of the paper involves both cluster-based and component-based subnetworks and searches for the inventors who bridge over the cluster boundary and thus enable the potential nurturing of biotechnology clusters with new (to the cluster) external knowledge. Since these inventors stand at the gate through which external knowledge enters clusters, we shall call them *gatekeepers* in the sense used in the literature surveyed above.

5.1 Three types of inventors

First we roughly categorize all Canadian inventors residing in the twelve studied geographical clusters based on each inventor's connections with other inventors. Three categories of inventors are established: internal, external and intermediary. An *internal inventor* only has intra-cluster connections, i.e. no collaboration partner outside the cluster. An *external inventor* does not participate in any intra-cluster collaboration, since all of his links are directed out of the cluster. Even if he physically resides in the cluster he has no contacts there and any external knowledge which he acquires remains on the cluster's border. None of the internal or external inventors can thus contribute to the actual knowledge transmission between clusters; an *intermediary* however maintains both intra-cluster and inter-cluster connections and as such, his existence is instrumental to the potential delivery of fresh outside knowledge to the cluster. Out of 3065 inventors residing in Canadian clusters, 31% (936 inventors) are such intermediaries.

The importance of an intermediary may be measured by the amount of knowledge he may provide to the cluster, for which the number of direct sources/inventors of external knowledge to which each intermediary is connected is a proxy. Table 4 shows the average number of inter-cluster links (or inter-lines, in the fourth column) for intermediaries in each cluster, which corresponds to the amount of knowledge an average intermediary potentially delivers to his cluster. Moreover, the third column displays the average number of links (or average degree), including both intra-cluster and inter-cluster, that are connected to the intermediaries in each cluster. This measure indicates how well an average intermediary is interconnected in general. Furthermore, we have grouped the intermediaries based on the number of their inter-cluster links, the results of which are provided in the last four columns of the same table. Around 70% of all intermediaries collaborate with only 1 or 2 out-of-cluster partners and are thus connected to only 1 or 2 channels through which they could introduce external knowledge into the cluster. An intermediary with a low number of external connections could still be extremely important for the cluster as a transmitter of external information, since this also depends on his position in the network.

(Insert Table 4 here)

5.2 From intermediaries to gatekeepers

In order to evaluate the positions of the intermediaries in the network we use the notion of *betweenness centrality***.** Since this measure does not distinguish between the place and direction of knowledge transmission (whether the inventor serves as an important intermediary mainly among the inventors from the same cluster or he is indeed instrumental in the external knowledge transfer to the inventors in the cluster), it cannot fully capture how strategic an inventor's position is as an external knowledge procurer.

At this point we thus use betweenness centrality merely to filter out intermediaries whose betweenness is zero, since any external knowledge transmitted through such inventors is redundant. For instance, imagine an inventor *i* connected to the same exact inventors as at least one other inventor *j* in the component (who is a co-author on all the same patents as *i* and hence potentially transmits exactly the same knowledge as the original inventor *i*). If inventor *j* has collaborated on a single additional patent without inventor *i*, then there is at least one other intermediary in the cluster which has exactly the same connections as the original inventor *i* plus at least one additional connection leading to other inventors. The obtained betweenness of the original inventor *i* will thus equal zero.

Betweenness centrality in fact measures how the disappearance of an inventor would alter the shortest paths and connectedness between all other inventors. Since the disappearance of inventors with zero betweenness would neither reduce the amount of external knowledge which potentially enters the cluster nor the speed at which it could enter (no shortest path would get longer), they are considered redundant and hence excluded from further analysis.

After this filtering process, only around half the intermediaries (434 or 14% of all Canadian inventors within clusters) are retained. Even though for the purpose of the analysis to follow, we do not consider the redundant intermediaries, they are nevertheless important in the regional system of innovation, as knowledge can possibly "enter" the cluster from a number of sources. Performing once again the interlines analysis exclusively for the non-redundant intermediaries yields Table 5 and allows a comparison with the previous results including all intermediaries (in Table 4).

(Insert Table 5 here)

The comparison suggests that most redundant intermediaries have a very low number of ties to external knowlegde sources as the percentage of intermediaries with only 1 or 2 connections outside the cluster dropped from around 70% to about 50%. This shows that non-redundant intermediaries are usually better interconnected with out-of-cluster collaborators. A proportionally much greater amount of non-redundant intermediaries with many direct sources of external information (6 or more inter-lines) is found in the clusters of Saskatoon (35%) and Calgary (25%), whereas in the big clusters of Toronto, Montreal and Vancouver, almost 90% of all outside knowledge is brought into the clusters by less connected non-redundant intermediaries (1-5 inter-lines). In fact, this is already detectable in the analysis of all intermediaries in Table 4, but the exclusion of the redundant gatekeepers made this observation more pronounced.

5.3 Important non-redundant intermediaries: the gatekeepers

Let us now examine only the 25 most important non-redundant intermediaries, i.e. those with the highest number of direct sources of outside knowledge and order them according to the number of their inter-cluster links (Table 6). One inventor from Toronto (**$\mathbf{TRT_1}$**) has the highest number of direct external sources (29). The sum of the value of all his inter-lines is 81, i.e. this inventor has collaborated with 29 external collaborators on 81 occasions. The next column shows the degree of a vertex (inventor), which is the sum of all his links, including both inter-cluster and intra-cluster. The inventor **$\mathbf{TRT_1}$** has only four additional links within the cluster (his degree is 33), which means that all the external knowledge that he acquires flows further into the cluster only through 4 of his colleagues from the cluster.

Since not all inventors in the clusters are interconnected within the cluster itself, we do not know how many of them benefit from the external knowledge introduced by any particular intermediary. These indicators do not allow the measurement of whether an inventor is alone in effectively transferring external knowledge to these inventors or whether there are others contributing to this task (which would make his contribution less critical). Moreover, we are not able to assess how much innovative potential this knowledge may create. As a consequence, we have developed several measures to help answer these questions. In order to evaluate the importance of each inventor for the capacity to transmit of external knowledge and to assess the external innovative potential delivered by him to other inventors in the cluster we have created a *Gatekeeper's Importance Index (GII)* both for the cluster and for Canada.

(Insert Table 6 here)

Let us first start with the definition necessary for understanding the concept: A *Cluster-Component group of inventors (C-C group)* is a group of inventors residing in a Canadian cluster who are all directly or indirectly interconnected within the cluster. In a great majority of components, the C-C groups were created as a simple intersection between the clusters and the components, however - particularly in the 4 largest components - many inventors residing in the same cluster and being part of the same component are not directly connected within the cluster and end up in different C-C groups. Figure 3 illustrates the position of the three types of inventors of Component C1.

(Insert Figure 3 here)

In the centre of the figure is the largest group of inventors in this component, which is composed mainly of foreigners but also of some Canadian inventors residing outside clusters. It is fairly obvious that it is these predominantly foreign inventors who are interconnecting all other Canadian inventors in this component. Many of the inventors within the component do not have any other connection among themselves except through foreign inventors. Canadian inventors located in clusters are depicted here in three concentric circles around the core of foreigners and out-of-cluster inventors. The inner circle is composed of external inventors, which do not have any “direct” connections with their fellow inventors from the cluster, but indirectly through out-of-cluster and foreign inventors. Each of these external inventors actually constitutes a separate C-C group (those formed by the external inventors are neither indicated in the figure nor discussed further). In the middle circle are located the inventors connected to those residing both outside and inside the cluster – these are the intermediaries. The rest of the inventors - placed in the outer circle (on the periphery of the figure) - are internal cluster inventors connected only to intermediaries or among themselves. Many inventors in the larger clusters had to be separated, notably in Toronto and Vancouver where they ended up in 5 different C-C groups in each cluster, since the only connections existing between them are through inventors residing outside clusters.

The *Gatekeeper’s Importance Indices (GIIs)* are based on the measurement of the importance of each intermediary as a potential source of external information for the C-C group to which he takes part and the importance of this C-C group either for the cluster or for Canada. The two *GIIs* are defined as:

$$GII_i^{cluster} = \frac{I_i}{I_{cc}} \cdot \frac{P_{cc}}{P_{cluster}} \cdot B_i \cdot 1000$$

$$GII_i^{Canada} = \frac{I_i}{I_{cc}} \cdot \frac{P_{cc}}{P_{Canada}} \cdot B_i \cdot 1000$$

where:

- $GII_i^{cluster}$ …*Gatekeeper's Importance Index for Cluster* for inventor *i*
- GII_i^{Canada} …*Gatekeeper's Importance Index for Canada* for inventor *i*
- I_i…the number of inter-cluster links of the inventor *i*
- I_{CC}...the sum of all inter-cluster links of the C-C group cc (which includes inventor *i*)
- P_{cc}…the sum of all the patents invented or co-invented by at least one inventor from the C-C group cc (which includes the inventor *i*)
- $P_{cluster}$…the sum of all the patents authored or co-authored by all the inventors in the cluster in which the inventor *i* resides
- P_{Canada}… the sum of all the patents authored or co-authored by all the inventors residing in Canadian clusters
- B_i…betweenness centrality of the inventor *i*

The first term of the product in both indices captures the importance of the inventor as a potential source of external information for the C-C group. It measures the number of inter-links connected to each inventor (I_i) as a share of all the inter-links entering the given C-C group of inventors (I_{CC}). Since we disregard time in this analysis and thus assume that all links are active simultaneously, we can also assume that the amount of external knowledge incoming by each such channel is equal whatever the values of the links. The values of the links might show the efficiency with which the information is exchanged but do not reveal anything about the total amount of information which could be transmitted through the particular channel. This remains to be the same no matter how many times the collaboration between the two inventors took place and depends solely on the availability of the knowledge sources of the inventor on the other side of the channel.

The second term of $GII_i^{cluster}$ evaluates the importance of each C-C group for the cluster based on the innovative productivity of that group. The patents which are authored or co-authored by at least one of the C-C group inventors are added for each group and divided by the sum of all the patents invented or co-invented by at least one of the inventors from the cluster ($P_{cluster}$). The last importance measure, which constitutes the second term of GII_i^{Canada} evaluates the importance of each C-C group for Canada and is based on the innovative productivity of the group as well. It also counts the number of patents that have been created within the C-C group of a given inventor and expresses that number as a share of the total innovative production in all Canadian clusters (P_{Canada}).

The last term of the product in both indices measures the betweenness centrality of the inventor (B_i) and indicates how well the inventor is interconnected in general[15]. This involves an overall evaluation of his network position which goes far beyond the external channels: it takes into consideration his other connections inside the cluster, the connections of all the inventors to whom he is connected and the positions of all the other inventors in the component from which he can indirectly gather knowledge or to whom he can deliver it.

The resulting products are called *Gatekeeper's Importance Indices* and measure an inventor's importance as a procurer of external knowledge for the cluster ($GII_i^{cluster}$) or for Canada (GII_i^{Canada}) based on the share of innovative production to which he thereby contributes.

---

[15] It is in part for the calculation of these indices that we ignore the redundant gatekeepers.

5.4 Canadian biotechnology gatekeepers

Table 6, which presents the importance measures for the 25 intermediaries with the highest number of direct external sources, contains all the importance indices as well. Here are few examples which show how to interpret the measures: inventor **TRT_1** has the greatest count of inter-cluster collaboration links and contributes to around 24% of all the potential external knowledge input flowing into his C-C group (i.e. the percentage of TRT_1 interlinks with respect to the total number of interlinks of the cluster). The C-C group's share of the patent production represents around 4% of the cluster's production and around 1.5% of the total Canadian patent production. The final *Gatekeeper's Importance Indices*, which also take into account his network position, place inventor **TRT_1** in 8th position for his importance in the cluster and in 12th position for his importance in Canada. Within his own Toronto cluster, he is the 4th most important inventor in terms of his function as an intermediary of external information.

Inventor **CAL_2** brings over 76% of external knowledge into the C-C group; this group however does not contribute significantly to the overall patent production in the cluster (2.4%) and even less in Canada (0.1%). Furthermore, even though **CAL_2** has 13 direct sources of information outside the cluster his C-C group inside the cluster is actually formed only by him and one additional inventor and his betweenness score is very low. In spite of the high number of external sources to which he has a direct access, the importance of this intermediary is quite negligible and he ranks very low both in his cluster and in Canada.

Similar situation can be observed for the inventors **TRT_1**, **OTT_2**, **KIN_1** and **TRT_7**. These intermediaries appear to utilize a relatively large number of direct sources of external information for themselves, but they do not transfer the knowledge to many fellow inventors inside their own clusters. It would seem that these gatekeepers act in fact as ambassadors of knowledge from their own clusters to the outside world, i.e. what Gould and Fernandez (1989) identify as representatives.

Four inventors with the highest scores of $GII_i^{cluster}$ in Canada are from the Saskatoon and Calgary clusters, which points out towards the crucial role played by these intermediaries in their own cluster. Table 7 presents the average importance indices for all inventors acting as intermediaries for the cluster (in the third column). It shows that the average scores of $GII_i^{cluster}$ for Calgary (0.04) and Saskatoon (0.03) are much higher than that of any other cluster. The situation changes slightly when the average importance indices for Canada (GII_i^{Canada}) are calculated (in the fifth column[16]). Inventors from Toronto significantly gain in importance as gatekeepers for Canada (10 out of the first 20 intermediaries with the highest GII_i^{Canada} are from Toronto).

(Insert Table 7 here)

6. Discussion

In comparing the cluster-based subnetworks with the component-based subnetworks, we showed the fragmentation of the former subnetworks. If there are structural holes between the components within a geographical cluster (what we refer to as the internal inventors), they are not generally bridged within its boundary unless the cluster possesses a critical mass of inventors.

We have already estabished that *internal* and *external* inventors do not participate in the transmission of external knowledge to the cluster, since they lack either the connection outside or

---

[16] A consequence of our extraction methodology, mentioned in footnote 12, is the lack of precision of the Gatekeeper's importance index for Canada. Because we do not have all patents, we cannot truly assess the betweenness centrality of Canadian gatekeepers within the world biotechnology network. This last column must therefore be interpreted in consequence.

inside their cluster. According to Figure 4, these inventors constitute the majority of inventors in all geographical clusters (60%-80% for most clusters). Inventors which do maintain both intra-cluster and inter-cluster collaborations, but do not serve as indispensible intermediaries for other inventors are *redundant intermediaries*. These inventors can still be productive and thus considered important creators of biotechnology innovation (even star scientists as described by Audretsch and Stephan, 1996; Zucker and Darby, 1996), but they are redundant as external information procurers. Around 15%-20% of inventors in most of the geographical clusters are such intermediaries.

(Insert Figure 4 here)

The remainder of the inventors are considered to be the *gatekeepers*. These are the intermediaries which potentially introduce non-redundant knowledge to the cluster and thereby contribute to the innovative potential of other inventors in the cluster. The highest percentage of gatekeepers among the cluster's inventors is found in Calgary (26%), Edmonton (20%) and Ottawa (20%), whereas Vancouver (9%) and the small clusters (6%-12%) have the lowest shares. However, the levels of contribution differ significantly among the gatekeepers themselves and therefore we have designated any gatekeeper with $GII_i^{cluster}$ of at least 0,001 as an *important gatekeeper*. Quite high percentages (around 60%) of all gatekeepers are considered to be important gatekeepers in the clusters of Saskatoon and Ottawa. In the greatest clusters of Toronto, Montreal and Vancouver however, only around 10%-13% of all gatekeepers are important gatekeepers for the cluster (the number of the important gatekeepers in Ottawa is higher than their count in Toronto even in absolute terms). Besides a possible size effect for the smaller clusters, we propose that the three main clusters possess enough of a critical mass of inventors so that the need for out-of-cluster knowledge is reduced and so is the need for a high proportion of very important gatekeepers. In other words, the structural holes can be found within the clusters. The relative contribution of the Toronto gatekeepers to the total Canadian biotechnology innovation production is however much more important.

Most of the network components (758 components, which represents 85% of all components) do not involve any gatekeeper. These are either components with only internal and external inventors (often single-inventor components or isolates) or components where all the inventors are connected to each other (each inventor is an intermediary who potentially absorbs outside knowledge, but does not transmit it any further, since all of his colleagues have access to the same knowledge sources, i.e. they are all redundant intermediaries.). As for the components with gatekeepers (136 components, or 15% of the total), over half of them involve only one gatekeeper for the entire component. In this case there is one C-C group within the component where all external knowledge could be transferred to the group only through a single intermediary. If there are any other C-C groups within such component they consist either only of an external inventor or only of redundant intermediaries. Almost half (44%) of the 434 gatekeepers are part of the four largest components. This highlights the critical role played by the large components in the introduction of new knowledge to the cluster.

Figure 3 illustrates the collaboration pattern among inventors within the largest component in the Canadian biotechnology network (Component C1, which involves 24% of all gatekeepers). It shows that inventors within the same cluster may not in fact be connected within the cluster and a foreign or out-of-cluster inventor is necessary to transmit knowledge between them. Within the same cluster and component there are groups working completely separately and the short geographical distance between them does not seem to play a role when seeking for collaboration partners.

This allows us to make some conjectures about the position of the Canadian biotechnology network in the worldwide biotechnology innovation network. Many Canadian inventors who now seem to be disconnected may in fact be part of the same international component in the worldwide biotechnology innovation network. The complete Canadian biotechnology network would then be in fact much less fragmented than we see it now and there may exist one giant Canadian biotechnology network component, which would comprise a great majority of inventors as suggested by Newman (2001a). Furthermore, if we extend this theory further, most biotechnology inventors in the world might in fact be united in one giant international component where they all indirectly collaborate, share their knowledge and create collective inventions.

7. Implications and direction for future research

Our goal in this paper was to develop a systematic approach to identify the inventors that every biotechnology firm would want to employ. As the old adage suggests, “it’s not what you know but who you know” that matters. These individuals are well connected, have access to a number of external sources of knowledge built over the years from numerous patent collaborations with inventors from their geographical cluster and beyond its boundary. As such, these gatekeepers span over the structural holes suggested by Burt (1992).

We have also shown that geographic proximity is not a universal panacea and as suggested by a number of scholars (for instance Boschma *et al.*, 2007; Wink, 2008), other types of proximity, are present and essential in the knowledge generation process. These studies have generally examined the position of firms in geographical clusters that are part of knowledge networks or not, to see weather geographic proximity is sufficient for knowledge transmission and adoption. We have dug deeper into firms, universities and other organisations to study the interaction between the cluster-based collaboration and the out-of-cluster collaboration leading to biotechnology patents. Because of the high mobility of scientists, inventors and engineers noted in the literature, interactions between organizations lose their precision and it seems more appropriate to then follow the individuals that are at the core of these knowledge exchanges.

In addition, concentrating research within clusters or other smaller geographical regions leaves a great deal of connections out of the picture. For instance, what may have been construed as localised knowledge spillovers, in fact may simply be the result of knowledge transfer via a foreign or out of cluster individual or entity. Furthermore, geographical proximity is not enough and “knowledge may be far from accessible to most of those who are located nearby its sources” (Breschi and Lissoni, 2001b, p. 262). Social proximity within epistemic communities that share a common jargon and a common knowledge-base offer a richer environment for knowledge diffusion. Studying one type of proximity and leaving the other misses much of the potential for knowledge diffusion.

Certain limitations in this study should be taken into consideration for future research. In this paper, we have measured the importance of various gatekeepers by the number of patents they invented or co-invented. As such, we only measure the potential knowledge transmission capability of a gatekeeper without specifically investigating what is really exchanged. Adding and comparing citation rates received by the patents co-invented by gatekeepers to this measure would allow the evaluation of which gatekeepers are actually efficiently transmitting knowledge to the geographical cluster and its impact on innovation. Frenken *et al.* (2005) for instance find that scientific articles co-authored by teams spanning distinct organizations and countries receive higher citation rates. Similarly to Powell *et al.* (1996) using all types of collaborative activities to build the network surrounding biotechnology firms, we plan to exploit the co-inventorship and citations for the construction of a multi-links network. Such a multi-level network will allow us to investigate if the position of an inventor in co-inventorship network has an influence on the number of citations gained by the individual or the cluster-component to which he is related. This will require a dynamic analysis of the network which we address below.

In addition to adding patent citations to our database, we are currently in the process of identifying the organization of each inventor by merging our biotechnology patent data with scientific articles data using the affiliations of each author is listed. This will allow a multi-level analysis in the spirit of Klein and Kozlowski (2000) where organization boundaries and work philosophies (firms versus universities for instance) could be explored. An individual would thus be part of an organization within a geographical cluster and part of a social network.

Another line of enquiry that we plan to follow therefore consists of adding a number of personal attributes (as suggested by Fleming *et al.* 2007), such as age, sex, experience, career path and star scientist status to the database. This would allow the investigation of whether the most productive inventors are also the best procurers of external knowledge for the cluster or whether there is a division of "labour" between the two inventors attributes. This process is very cumbersome but should allow us to evaluate whether university-inventors more often act as gatekeepers than their industrial counterparts.

Throughout the paper, we have assumed as shown by Dahl and Pedersen (2005) that once individuals collaborate formally, they remain in contact for a number of years. Very few individuals in our database are however present for more than 10 years. Because of the lengthy process of the development of human biotechnology innovation (roughly 10 years), it is not inconceivable that inventors remain in contact for such a long period of time. This does not allow us to examine how one becomes or remains a gatekeeper, how inventor mobility or affiliation facilitates gatekeeping activities, and so on. We therefore plan to perform the analysis using 5, 10 and 15 years time windows to gain some insights on gatekeeping dynamics. Because of the lengthy time necessary to develop biotechnology products and processes, shorter time periods as used by Powell *et al.* (1996) would not be appropriate. Our methodology allows the identification of the individuals that potentially have the most impact on bringing fresh knowledge to a cluster. As such, surveying a number of these individuals to try to understand how they ended up in these gatekeeping positions is the next step of this research.

In this paper, the focus was on Canada. It would be interesting to compare the results obtained with other countries to see whether the proportion of gatekeepers differs. Casper and Murray (2005) for instance find that British scientists are much more mobile, from big pharmaceutical companies to biotechnology start-up firms, than their German colleagues. This facilitated mobility contributes to enlarging their social network which in turn may improve their positions as gatekeepers. Finally, as suggested above, an addition of all the worldwide biotechnology patents would allow to see the networks in their entirety and to gain a full picture of innovation production in Canadian biotechnology as compared with other countries. This paper is thus a first step towards the understanding of the role and importance of gatekeepers.

8. Conclusions

This paper studies the social networks of inventors in which a tie between two actors represents a co-inventorship of one or more patents. Drawing from the list of inventions from the USPTO website, we have created a patent database and constructed the innovation network for all registered biotechnology patents in which at least one inventor or co-inventor resides in Canada. We have examined the structure of the collaborative networks within two different concepts: First, collaboration among the inventors working in close geographical proximity – within geographical clusters; second, collaborative ties among the inventors who are directly or indirectly interconnected in network components disregarding geographical distances.

We find that the cluster-based subnetworks and the components (or component-based subnetworks) overlap to a certain extent, but differ in their structure. Many inventors from the same geographical cluster may also be part of the same network component. The bulk of smaller components are entirely contained within one cluster, larger components however usually span over several clusters. Moreover, most of the larger or medium-sized components include foreign collaborative relations as well. We find that these foreign inventors are extremely important in connecting Canadian inventors from different clusters together (or even from the same cluster - particularly in the largest components), which makes their presence critical for the transmission of knowledge between Canadian inventors. We conjectured that if all biotechnology patents in the world were included in the analysis, the Canadian biotechnology network would be less fragmented and most of the inventors would in fact be a part of one giant international biotechnology innovation component in which all inventors indirectly collaborate, share their knowledge and create collective inventions.

We also investigate the points of interaction at the frontier of the geographical clusters. In order to understand exactly how knowledge travels among clusters through the channels of network components, we have searched for gatekeepers – the inventors who bridge over the two spaces and thus potentially facilitate the nurturing of biotechnology clusters with fresh external knowledge. In order to systematically identify these gatekeepers, we have created two indicators, which measure each inventor’s importance as a procurer of external knowledge for the cluster and for Canada, based on the share of innovative production to which he thereby contributes. Only around 10%-20% of all inventors in most clusters are identified as gatekeepers and are responsible for the inflow of external information to the cluster. These inventors are nevertheless crucial to the innovation process and are priceless commodities for a firm.

Although we agree that further research is required to identify how one becomes a gatekeeper, what characteristics must a gatekeeper have and what are the mechanisms used for knowledge transmission, our paper nevertheless provides a systematic approach to identify the most important gatekeepers. These individuals are important even more so for small clusters that aim to grow and stay innovative by keeping the door open to the outside world. Our approach provides a means to identify the important individuals of a network with distant connections in order to further investigate the mechanisms of gatekeeping or knowledge transmission over long distances.

References

Allen, T.J., 1967. Communication in the research and development laboratory. Technology Review, 70 (1): 31-7.

Allen T.J. ,1977. Managing the Flow of Technology. Technology Transfer and the Dissemination of Technological Information within the R&D Organization. MIT Press, Cambridge.

Audretsch, D.B. and Feldman, M.P., 1996. R&D spillovers and the geography of innovation and production. The American Economic Review, 86(3): 630-640.

Audretsch, D., Stephan, P., 1996. Company-scientists locational links: The case of biotechnology. American Economic Review, 86: 641-652.

Baptista, R. and Swann, P., 1998. Do firms in clusters innovate more? Research Policy, 27: 525-540.

Beaudry, C., 2001. Entry, growth and patenting in industrial clusters. International Journal of Economics of Business, 8(3): 405-436.

Beaudry, C. and Breschi, S., 2003. Are firms in clusters really more innovative? Economics of Innovation and New Technology, 12(4): 325-342.

Beaudry, C., Schiffauerova, A., 2009 forthcoming. Biotechnology and Nanotechnology Innovation Networks in Canadian Clusters, in Laperche, B. (ed.) Innovation Networks and Clusters,Routledge.

Boschma, R.A., 2005. Proximity and innovation: A critical assessment. Regional Studies, 39(1): 61-74.

Boschma R.A. and Ter Wal, A.L. J., 2007. Knowledge networks and innovative performance in an industrial district: The case of a footwear district in the South of Italy. Industry and Innovation, 14(2): 177.

Breschi, S. and Lissoni, F., 2001a. Knowledge spillovers and local innovation systems: A critical survey. Industrial and Corporate Change, 10(4): 975-1005.

Breschi, S. and Lissoni, F., 2001b. Localized knowledge spillovers vs. innovative milieux: Knowledge "tacitness" reconsidered. Papers in Regional Science, 80:255-273.

Burt, R.S., 1992. Structural Holes: The Social Structure of Competition. Harvard University Press: Cambridge, MA.

Burt, R., 2001. Bandwidth and echo: Trust, information, and gossip in social networks. In: A. Casella and J.E. Rauch (Eds.), Networks and Markets: Contributions from Economics and Sociology. New York: Russel Sage Foundation.

Burt, R.S., 2004. Structural holes and good ideas. American Journal of Sociology, 110(2): 349-399.

Burt, R.S., 2007. Secondhand brokerage: Evidence on the importance of local structure for managers, bankers, and analysts. Academy of Management Journal, 50(1): 119-148.

Casper, S. and Murray, F., 2005. Careers and clusters: analyzing the career network dynamic of biotechnology clusters. Journal of Engineering and Technology Management, 22(1-2): 51-74.

Cowan, R., David, P.A., and Foray, D., 2000. The explicit economics of knowledge codification and tacitness. Industrial and Corporate Change, 9(2): 211-253.

Cowan, R. and Jonard, N., 2004. Network structure and the diffusion of knowledge. Journal of Economic Dynamics and Control, 28: 1557-1575.

Dahl, M. S., Pedersen, C. Ø. R., 2005. Social networks in the R&D process: the case of the wireless communication industry around Aalborg, Denmark. Journal of Engineering and Technology Management, 22(1-2): 75-92.

de Nooy, W., Mrvar, A. and Batagelj, A., 2005. Exploratory Social Network Analysis with Pajek. Cambridge: Cambridge University Press.

Fernandez, R. M. and Gould, R.V., 1994. A dilemma of state power: Brokerage and influence in the national health policy domain. American Journal of Sociology, 99: 1455-1491.

Fleming, L., King, C.III. and Juda, A., 2006. Small worlds and regional innovation. SSRN Working Papers.

Fleming, L., Mingo, S. and Chen, D., 2007. Collaborative brokerage, generative creativity, and creative success. Administrative Science Quarterly, 52(3): 443.

Frenken, K., Hölzl, W. and de Vor, F., 2005. The citation impact of research collaborations: the case of European biotechnology and applied microbiology (1988–2002). Journal of Engineering and Technology Management, 22(1-2): 9-30.

Gould, R.V. and Fernandez, R.M., 1989. Structures of mediation: A formal approach to brokerage in transaction networks. Sociological Methodology, 19: 89-126.

Graf, H., 2008. Gatekeepers in regional networks of innovators. The 25th DRUID Celebration Conference on Entrepreneurship and Innovation - Organizations, Institutions, Systems and Regions, Copenhagen, CBS, Denmark.

Hargadon, A. and Sutton R.I., 1997. Technology brokering and innovation in a product development firm. Administrative Science Quarterly, 42: 716-749.

Howells, J., 2006. Intermediation and the role of intermediaries in innovation. Research Policy, 35: 715–728.

Jaffe, A.B. and Trajtenberg, M., 1996. Flows of knowledge from universities and federal laboratories. Modeling the flow of patent citations over time and across institutional and geographic boundaries. Proceedings of the National Academy of Sciences, 93.

Katz, R. and Tushman, M.L., 1981. An investigation into the managerial roles and career paths of gatekeepers and project supervisions in a major R&D facility. R&D Management, 11: 103-110.

Kirat, T. and Lung, Y., 1999. Innovation and proximity. Territories as loci of collective learning processes. European Urban and Regional Studies, 6(1): 27-38.

Kozlowski, S., & Klein, K. J. 2000. A multilevel approach to theory and research in organizations: Contextual, temporal and emergent processes. In K. J. Klein, & S. Kozlowski (Eds.), Multilevel theory, research, and methods in organizations: Foundations, extensions and new directions: 3-90. San Francisco, CA: Jossey-Bass.

Lazaric, N., Longhi, C. and Thomas., C. 2008. Gatekeepers of knowledge versus platforms of knowledge: From potential to realized absorptive capacity. Regional Studies, 42(6): 837.

Lissoni, F., 2001. Knowledge codification and the geography of innovation: The case of Brescia mechanical cluster. Research Policy, 30: 1479-1500.

Malipiero, A., Munari, F. and Sobrero, M., 2005. Focal firms as technological gatekeepers within industrial districts: Knowledge creation and dissemination in the Italian packaging machinery industry. DRUID Working Paper No. 05-05

Marsden, P.V., 1982. Brokerage behavior in restricted exchange networks. In: Marsden, P. and Lin, N. (Eds.), Social Structure and Network Analysis, Beverly Hills: Sage.

Morrison, A., 2008. Gatekeepers of knowledge within industrial districts: Who they are, how they interact. Regional Studies, 42(6): 817.

Newman, M.E.J., 2001. Scientific collaboration networks. I. Network construction and fundamental results. Physical Review, 64, 016131.

Nooteboom B., 1999. Innovation, learning and industrial organization. Cambridge Journal of Economics, 23, 127–150.

Owen-Smith J. and Powell W.W., 2004. Knowledge networks as channels and conduits: The effects of spillovers in the Boston biotechnology community. Organization Science, 15(1): 5-21.

Pandza, K. and Holt, R., 2007. Absorptive and transformative capacities in nanotechnology innovation systems. Journal of Engineering and Technology Management, 24(4): 347-365.

Powell, W. W., Koput, K. W. and Smith-Doerr, L., 1996. Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. Administrative Science Quarterly, 41(1): 116-145.

Schiffauerova, A. and Beaudry, C., 2009. Geographical aspects of collaborative networks in Canadian biotechnology clusters, in Nolin, T. P., Handbook of Regional Economics, Nova Publishers., pp.193-221

Sorenson, O., 2003. Social networks and industrial geography. Journal of Evolutionary Economics, 13(5): 513-527.

Sorenson, O. and Stuart, T.E., 2001. Syndication networks and the spatial distribution of venture capital investments. American Journal of Sociology, 106(6): 1546-88.

Steiner, M. and Ploder, M., 2008. Structure and strategy within heterogeneity: Multiple dimensions of regional networking. Regional Studies, 42(6): 793.

Stuart, T.E. and Ozdemir, S.Z. and Ding, W.W., 2007. Vertical alliance networks: The case of university–biotechnology–pharmaceutical alliance chains. Research Policy 36: 477–498.

Torre, A. and Gilly, J-P., 2000. On the analytical dimension of proximity dynamics. Regional Studies, 34(2): 169-180.

Tushman, M.L. and Katz, R., 1980. External communication and project performance: An investigation into the role of gatekeepers. Management Science, 26: 1071-1085.

Waguespack, D. M. and Birnir, J. K., 2005. Foreignness and the diffusion of ideas. Journal of Engineering and Technology Management, 22(1-2): 31-50.

Wasserman, S. and Faust, K., 1994. Social Network Analysis, Cambridge: Cambridge University Press.

Winch, G. M. and Courtney, R., 2007. The organization of innovation brokers: An international review. Technology Analysis and Strategic Management, 19(6): 747-763.

Wink, R., 2008. Gatekeepers and proximity in science-driven sectors in Europe and Asia: The case of human embryonic stem cell research. Regional Studies, 42(6): 777-791.

Zucker, L. M., Darby, M., 1996. Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. Proceedings of the National Academy of Science, 93: 12709-12716.

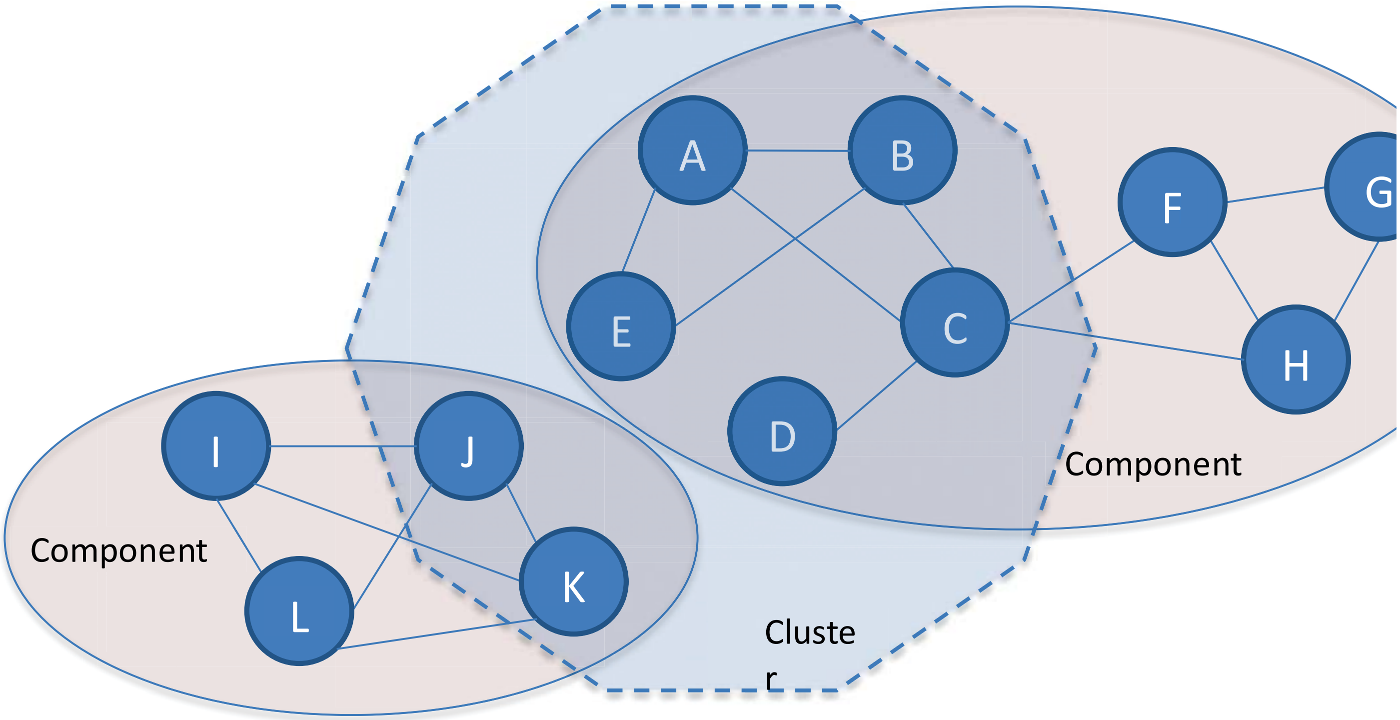


Figure 1[17]**:** Example of network composed of two components (component-based subnetworks)

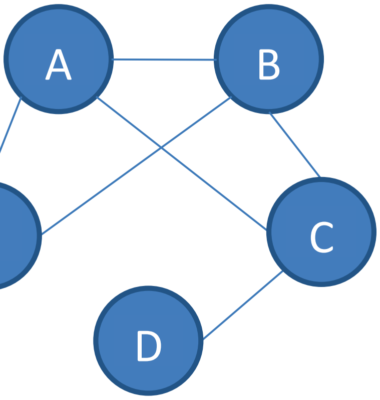


Figure 2: Network example where inventors A, B and C are co-inventors on a patent, inventors A, B and E are co-inventors on another patent and inventors C and D are co-inventors on a third patent. This is the part of one of the component located in the cluster-based subnetwork of **Figure 1**.

---

[17] There are two component-based subnetworks (Inventors A to H belong to one component and inventors I to L belong to the second component) and one cluster-based subnetwork (which includes all the inventors within the cluster, namely A to E, J, and K). Inventors F, G, H, I and L are located outside the cluster.

Table 1: Structural properties of the cluster-based subnetworks

Cluster[1]	**Canada**	**TRT**	**MTL**	**VAN**	**EDM**	**CAL**	**SAS**
Number of inventors	4569	927	698	411	210	91	147
Number of patents[2]	2485[3]	834	466	255	153	127	98
Number of collaborating pairs	9731	1120	1027	568	334	91	259
% of repeated collaborations	36%	43%	36%	37%	37%	41%	28%
Max number of repeated collaborations	60	60	11	10	14	16	8
STRUCTURAL COHESION							
Subnetwork density	0.001	0.003	0.004	0.007	0.015	0.022	0.024
Average degree	4.26	2.42	2.94	2.76	3.18	2.00	3.52
CENTRALIZATION OF SUBNETWORK							
Degree centralization	0.01	0.05	0.02	0.06	0.08	0.11	0.15
Betweenness centralization	0.009	0.008	0.011	0.005	0.019	0.011	0.074
CENTRALITY OF VERTICES							
Max degree centrality	66	51	16	27	20	12	25
Max betweenness centrality	0.009	0.008	0.011	0.005	0.019	0.011	0.076
GEODESIC DISTANCES							
Subnetwork diameter	17	9	11	5	7	5	6
Max reach	578	97	108	37	48	14	53
CLIQUISHNESS							
Average egocentric density	0.71	0.44	0.56	0.57	0.55	0.29	0.64
FRAGMENTATION							
# of components	894	342	218	134	67	39	34
Size of the 1st largest as % of all	13%	11%	16%	9%	23%	16%	37%
Share of components formed by 50% of inventors	10%	13%	11%	15%	11%	18%	6%
Isolates as % of inventors	4%	19%	15%	16%	17%	24%	13%

Cluster[1]	**WIN**	**KIN**	**OTT**	**QUE**	**HAL**	**SHE**
Number of inventors	77	94	224	127	33	26
Number of patents[2]	33	63	279	57	20	16
Number of collaborating pairs	54	96	343	155	20	10
% of repeated collaborations	19%	33%	36%	18%	50%	20%
Max number of repeated collaborations	3	10	19	7	5	3
STRUCTURAL COHESION						
Subnetwork density	0.018	0.022	0.014	0.019	0.038	0.031
Average degree	1.40	2.04	3.06	2.44	1.21	0.77
CENTRALIZATION OF SUBNETWORK						
Degree centralization	0.06	0.04	0.06	0.04	0.13	0.05
Betweenness centralization	0.002	0.003	0.068	0.003	0.010	0.000
CENTRALITY OF VERTICES						
Max degree centrality	6	6	16	8	5	2
Max betweenness centrality	0.002	0.003	0.070	0.003	0.010	0.000
GEODESIC DISTANCES						
Subnetwork diameter	3	3	11	4	2	1
Max reach	6	7	74	10	5	2
CLIQUISHNESS						
Average egocentric density	0.32	0.47	0.59	0.55	0.24	0.23
FRAGMENTATION						
# of components	44	38	70	44	20	18
Size of the 1st largest as % of all	9%	9%	33%	9%	18%	12%
Share of components formed by 50% of inventors	25%	24%	9%	21%	30%	33%
Isolates as % of inventors	36%	18%	17%	15%	39%	46%

[1] TRT ...Toronto MTL …Montreal VAN …Vancouver EDM …Edmonton CAL …Calgary SAS …Saskatoon
WIN …Winnipeg KIN …Kingston OTT …Ottawa QUE …Quebec HAL …Halifax SHE …Sherbrooke

[2] The numbers are based on the residence of the assignees and only the patents with at least one Canadian assignee are thus included

[3] Also includes the patents assigned outside the clusters or co-assigned to the several clusters at the same time

Table 2: Main characteristics and composition of the 30 largest components in the Canadian biotechnology innovation network

Component #	***C1***	***C2***	***C3***	***C4***	***C5***	***C6***	***C7***	***C8***	***C9***	***C10***	***C11***	***C12***	***C13***	***C14***	***C15***
# of inventors	579	185	175	78	50	44	39	36	30	29	27	27	27	24	23
# of patents	606	155	139	70	32	70	31	50	30	6	12	15	65	1	12
Patents/inventor	1.05	0.84	0.79	0.9	0.64	1.59	0.79	1.39	1.00	0.21	0.44	0.56	2.41	0.04	0.52
# of coll.[a] pairs	2057	560	517	185	167	105	83	92	336	89	61	44	46	276	87
% of repeated coll.	45%	40%	29%	38%	46%	43%	40%	82%	82%	48%	16%	16%	54%	100%	62%
Max # of repeated coll.	60	11	19	9	8	9	3	12	6	6	2	2	32	1	5
Number of the component's inventors in each cluster															
Toronto	154	16	8	16		35		5	22	1		25	22		1
Montreal	13	2	112	35	4		34								8
Vancouver	55	10	2		38			1			11		1		
Edmonton	50	1						2	1					1	
Calgary	20							9			3				
Saskatoon	54	40							2						
Winnipeg	1			7											
Kingston	9														
Ottawa	17	77	1	6											1
Quebec	9	2	1	1											1
Halifax		1						1							
Sherbrooke	2	2													
out-of-cluster	25	5		4								1			
abroad	170	29	51	9	8	9	5	18	5	18	13	1	4	23	12

Component #	**C16**	**C17**	**C18**	**C19**	**C20**	**C21**	**C22**	**C23**	**C24**	**C25**	**C26**	**C27**	**C28**	**C29**	**C30**
# of inventors	23	22	20	19	18	18	18	18	17	17	17	17	16	15	15
# of patents	16	10	10	37	7	14	10	8	13	7	8	12	22	5	8
Patents/inventor	0.7	0.45	0.5	1.95	0.39	0.78	0.56	0.44	0.76	0.41	0.47	0.76	1.38	0.33	0.53
# of coll. pairs	53	43	46	46	41	37	39	38	40	53	54	62	41	40	48
% of repeated coll.	34%	7%	9%	59%	7%	70%	23%	0%	18%	32%	67%	19%	24%	5%	88%
Max # of repeated coll.	3	3	3	26	2	3	3	1	5	2	5	12	12	2	4
Number of the component's inventors in each cluster															
Toronto	8		1	18	1	2	1	6				1	15		
Montreal	5	19	12			7	12				12				1
Vancouver					3	5				1					
Edmonton					1				10						
Calgary						1			1					2	
Winnipeg										1					
Kingston					4				1						5
Ottawa												7			
Quebec	1														
Halifax									1						
out-of-cluster	2		1						1	1					
abroad	7	3	6	1	9	3	5	12	3	14	5	9	1	13	9

Note: [a] *coll.* is short for *collaboration*

Table 3: Structural properties of the component-based subnetworks (see Appendix for explanation description of the structural properties)

Component Number	***C1***	***C2***	***C3***	***C4***	***C5***	***C6***	***C7***	***C8***	***C9***	***C10***	***C11***	***C12***	***C13***	***C14***	***C15***
Number of inventors	579	175	185	78	50	44	39	36	29	30	27	27	27	24	23
STRUCTURAL COHESION															
Subnetwork density	0.01	0.04	0.03	0.06	0.14	0.11	0.11	0.15	0.83	0.20	0.17	0.13	0.13	1.00	0.34
Average degree	7.11	6.40	5.59	4.74	6.68	4.77	4.26	5.11	23.17	5.93	4.52	3.26	3.41	23.00	7.57
CENTRALIZATION OF SUBNETWORK															
Degree centralization	0.10	0.10	0.10	0.12	0.45	0.37	0.24	0.27	0.19	0.26	0.27	0.16	0.48	0.00	0.27
Betweenness centralization	0.55	0.36	0.49	0.43	0.53	0.53	0.54	0.47	0.01	0.48	0.45	0.51	0.77	0.00	0.57
CENTRALITY OF VERTICES															
Max degree centrality	66	24	24	14	28	20	13	14	28	13	11	7	15	23	13
Max betweenness centrality	0.56	0.38	0.51	0.48	0.56	0.56	0.59	0.51	0.02	0.52	0.51	0.60	0.81	0.00	0.59
GEODESIC DISTANCES															
Component diameter	17	12	14	10	7	6	7	7	2	5	6	8	6	1	3
Average distance	7.09	5.31	6.58	4.78	2.75	3.10	3.18	3.10	1.17	2.57	2.90	3.65	2.68	1.00	1.98
Maximum reach	57	174	184	77	49	43	38	35	28	29	26	26	26	23	22
CLIQUISHNESS															
Average egocentric density	0.80	0.81	0.79	0.75	0.80	0.57	0.82	0.77	0.94	0.80	0.83	0.69	0.69	1.00	0.94

Component number	**C16**	**C17**	**C18**	**C19**	**C20**	**C21**	**C22**	**C23**	**C24**	**C25**	**C26**	**C27**	**C28**	**C29**	**C30**
Number of inventors	23	22	20	19	18	18	18	18	17	17	17	17	16	15	15
STRUCTURAL COHESION															
Subnetwork density	0.21	0.19	0.24	0.27	0.27	0.24	0.25	0.25	0.29	0.39	0.40	0.46	0.34	0.38	0.46
Average degree	4.61	3.91	4.60	4.84	4.56	4.11	4.33	4.22	4.71	6.24	6.35	7.29	5.13	5.33	6.40
CENTRALIZATION OF SUBNETWORK															
Degree centralization	0.32	0.42	0.67	0.63	0.49	0.26	0.57	0.85	0.59	0.34	0.54	0.62	0.37	0.71	0.30
Betweenness centralization	0.58	0.62	0.71	0.67	0.68	0.44	0.70	0.85	0.69	0.50	0.36	0.50	0.42	0.71	0.50
CENTRALITY OF VERTICES															
Max degree centrality	11	12	16	15	12	8	13	17	13	11	14	16	10	14	10
Max betweenness centrality	0.63	0.68	0.73	0.69	0.71	0.53	0.72	0.85	0.71	0.53	0.38	0.51	0.48	0.71	0.53
GEODESIC DISTANCES															
Component diameter	4	5	3	3	5	7	3	2	4	3	3	2	4	2	3
Average distance	2.49	2.76	1.96	1.89	2.17	2.89	2.01	1.75	1.95	1.93	1.68	1.54	2.20	1.62	1.85
Maximum reach	22	21	19	18	17	17	17	17	16	16	16	16	15	14	14
CLIQUISHNESS															
Average egocentric density	0.82	0.82	0.84	0.83	0.75	0.68	0.83	0.90	0.88	0.85	0.85	0.94	0.66	0.95	0.91

Table 4: Inter-lines analysis for all intermediaries

	Number of intermediaries	*as % of all*	*Average degree*	*Average number of interlines*	*Number of intermediaries with:*			
					1-2 inter-lines	*3-5 inter-lines*	*6-9 inter-lines*	*10 or more*
Toronto	247	27%	6.5	2.4	**187 (76%)**	44 (18%)	9 (4%)	7 (3%)
Montreal	244	35%	6.0	2.3	**174 (71%)**	53 (22%)	15 (6%)	2 (1%)
Vancouver	101	25%	5.7	2.2	**79 (78%)**	11 (11%)	6 (6%)	5 (5%)
Edmonton	92	44%	**7.3**	2.7	52 (57%)	27 (29%)	**13 (14%)**	
Calgary	35	38%	5.9	**3.3**	21 (60%)	7 (20%)	**5 (14%)**	**2 (6%)**
Saskatoon	36	24%	**8.8**	**3.2**	24 (67%)	4 (11%)	**6 (17%)**	**2 (6%)**
Winnipeg	27	35%	3.9	1.8	24 (89%)	3 (11%)		
Kingston	20	21%	4.8	2.4	13 (65%)	6 (30%)	1 (5%)	
Ottawa	97	43%	6.9	2.7	60 (62%)	27 (28%)	8 (8%)	2 (2%)
Quebec	29	23%	4.6	1.6	25 (86%)	4 (14%)		
Halifax	5	15%	5.0	1.8	3 (60%)	2 (40%)		
Sherbrooke	3	12%	3.3	2.0	2 (67%)	1 (33%)		
ALL	**936**	**31%**	**6.3**	**2.4**	**664 (71%)**	**189 (20%)**	**63 (7%)**	**20 (2%)**

Table 5: Inter-line analysis for non-redundant intermediaries only

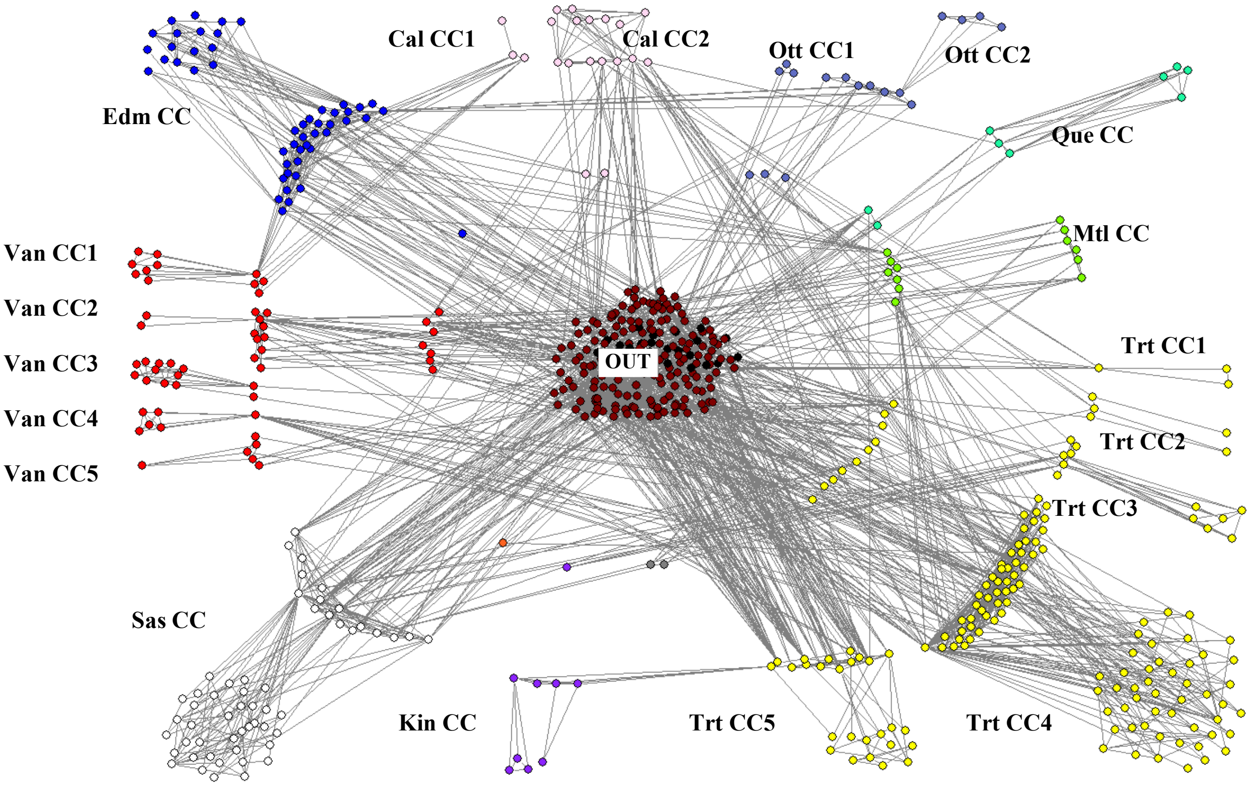
	Number of intermediaries	*As % of all*	*Average degree*	*Average number of interlines*	*Number of intermediaries with:*			
					1-2 inter-lines	*3-5 inter-lines*	*6-9 inter-lines*	*10 and more*
Toronto	124	13%	9.0	3.2	**74 (60%)**	**36 (29%)**	8 (6%)	6 (5%)
Montreal	111	16%	8.1	3.0	**56 (50%)**	**42 (38%)**	11 (10%)	2 (2%)
Vancouver	39	9%	7.7	2.6	**26 (67%)**	**8 (21%)**	3 (8%)	2 (5%)
Edmonton	41	**20%**	9.9	3.1	17 (41%)	18 (44%)	6 (15%)	
Calgary	24	**26%**	7.1	**4.0**	12 (50%)	6 (25%)	**4 (17%)**	**2 (8%)**
Saskatoon	20	14%	**12.6**	**4.6**	9 (45%)	4 (2%)	**5 (25%)**	**2 (10%)**
Winnipeg	5	6%	4.6	1.8	4 (80%)	1 (2%)		
Kingston	10	11%	5.9	2.6	6 (60%)	3 (3%)	1 (10%)	
Ottawa	45	20%	9.5	3.4	21 (47%)	16 (36%)	6 (13%)	2 (4%)
Quebec	9	7%	7.1	1.9	6 (67%)	3 (33%)		
Halifax	3	9%	5.7	2.3	1 (33%)	2 (67%)		
Sherbrooke	3	12%	3.3	2.0	2 (67%)	1 (33%)		
ALL	**434**	**14%**	**8.6**	**3.2**	**234 (54%)**	**140 (32%)**	**44 (10%)**	**16 (4%)**

Table 6: First 25 non-redundant intermediaries with highest inter-clines count showing values of all importance indices and other network properties

*Gate-keeper ID**	*Rank by inter lines*	*Inter lines count*	*Inter lines value*	*Degree*	*Between-ness*	*Patents by the C-C*	*C-C size*	*Importance of the inventor for C-C*	*Importance of C-C for cluster*	*Importance of C-C for Canada*	***Gatekeeper`s Importance Index***				
											$(GII_i^{cluster})$			(GII_i^{Canada})	
		(I_i)			*(1000xB$_i$)*	(P_{cc})		(I_i/I_{cc})	$(P_{cc}/P_{cluster})$	(P_{cc}/P_{Canada})	*Index value*	*Rank in cluster*	*Rank in Canada*	*Index value*	*Rank in Canada*
TRT$_1$	***1***	29	81	33	1.63	55	25	24.37%	4.05%	1.55%	0.0161	***4th in TRT***	***12***	0.0062	***8***
TRT$_2$	***2***	21	54	28	2.71	55	25	17.65%	4.05%	1.55%	0.0194	***3rd in TRT***	***11***	0.0074	***6***
CAL$_1$	***3***	16	94	25	8.94	64	17	33.33%	30.48%	1.80%	0.9085	***1st in CAL***	***1***	0.0537	***1***
TRT$_3$	***4***	15	59	66	7.21	254	110	8.62%	18.72%	7.15%	0.1163	***1st in TRT***	***5***	0.0444	***2***
MTL$_1$	***5***	14	27	19	0.39	117	112	9.09%	15.70%	3.30%	0.0056	***1st in MTL***	***33***	0.0012	***21***
CAL$_2$	***6***	13	14	14	0.01	5	2	76.47%	2.38%	0.14%	0.0001	***11th in CAL***	***155***	0.0000	***195***
TRT$_4$	***7***	12	38	28	0.66	254	110	6.90%	18.72%	7.15%	0.0085	***8th in TRT***	***23***	0.0032	***14***
TRT$_5$	***7***	12	12	17	0.01	8	6	70.59%	0.59%	0.23%	0.0000	***49th in TRT***	***195***	0.0000	***154***
MTL$_2$	***9***	11	11	19	0.13	117	112	7.14%	15.70%	3.30%	0.0014	***9th in MTL***	***68***	0.0003	***51***
SAS$_1$	***9***	11	21	33	3.30	80	54	13.10%	51.28%	2.25%	0.2219	***2nd in SAS***	***3***	0.0097	***4***
SAS$_2$	***9***	11	16	36	5.13	80	54	13.10%	51.28%	2.25%	0.3443	***1st in SAS***	***2***	0.0151	***3***
TRT$_6$	***9***	11	23	17	5.30	55	25	9.24%	4.05%	1.55%	0.0199	***2nd in TRT***	***10***	0.0076	***5***
VAN$_1$	***9***	11	12	22	1.00	8	12	42.31%	2.00%	0.23%	0.0085	***2nd in VAN***	***22***	0.0010	***22***
OTT$_1$	***14***	10	15	14	0.10	18	6	43.48%	5.94%	0.51%	0.0025	***11th in OTT***	***52***	0.0002	***70***
OTT$_2$	***14***	10	59	16	0.01	13	7	34.48%	4.29%	0.37%	0.0001	***29th in OTT***	***169***	0.0000	***191***
VAN$_2$	***14***	10	25	15	0.27	7	6	100.00%	1.75%	0.20%	0.0048	***4th in VAN***	***35***	0.0005	***35***
CAL$_3$	***17***	9	14	21	2.06	64	17	18.75%	30.48%	1.80%	0.1179	***2nd CAL***	***4***	0.0070	***7***
KIN$_1$	***17***	9	9	12	0.01	4	4	56.25%	4.12%	0.11%	0.0002	***4th in KIN***	***129***	0.0000	***205***
MTL$_3$	***17***	9	13	24	0.38	117	112	5.84%	15.70%	3.30%	0.0035	***3rd in MTL***	***42***	0.0007	***30***
SAS$_3$	***17***	9	38	16	0.02	80	54	10.71%	51.28%	2.25%	0.0009	***9th in SAS***	***85***	0.0000	***123***
SAS$_4$	***17***	9	46	16	0.02	80	54	10.71%	51.28%	2.25%	0.0009	***11th in SAS***	***85***	0.0000	***125***
SAS$_5$	***17***	9	38	19	0.05	80	54	10.71%	51.28%	2.25%	0.0025	***6th in SAS***	***51***	0.0001	***85***
SAS$_6$	***17***	9	38	16	0.02	80	54	10.71%	51.28%	2.25%	0.0009	***10th in SAS***	***85***	0.0000	***123***
TRT$_7$	***17***	9	58	12	0.01	15	5	60.00%	1.11%	0.42%	0.0001	***43rd in TRT***	***176***	0.0000	***135***
TRT$_8$	***17***	9	9	17	0.44	254	110	5.17%	18.72%	7.15%	0.0043	***11th in TRT***	***39***	0.0016	***18***

* Gatekeeper ID is based on the cluster of the inventor's residence and his rank according to the number of inter-lines. (Whereas the ranking in the 12th column is based on the values of $GII_i^{cluster}$):

TRT$_\#$	…Inventor of rank # in Toronto	MTL$_\#$	…Inventor of rank # in Montreal	VAN$_\#$	…Inventor of rank # in Vancouver
KIN$_\#$	…Inventor of rank # in Kingston	CAL$_\#$	…Inventor of rank # in Calgary	SAS$_\#$	…Inventor of rank # in Saskatoon
OTT$_\#$	…Inventor of rank # in Ottawa				



The vertices of different shades of grey indicate the inventors residing in different clusters.

Edm CC	…Edmonton C-C group	Mtl CC	…Montreal C-C group
Van CC#	…Vancouver C-C groups	Que CC	…Quebec C-C group
Sas CC	…Saskatoon C-C group	Ott CC#	…Ottawa C-C groups
Kin CC	…Kingston C-C group	Ca CC#	…Calgary C-C groups
Trt CC#	…Toronto C-C groups	OUT	…foreigners or Canadians outside clusters

Figure 3: Component C1 with all created C-C groups

Table 7: Average values of the indices of importance for the gatekeepers in each cluster

Cluster	*Importance of intermediary for the C-C*	*Importance of the C-C for cluster*	***Gatekeeper's importance index for cluster***	*Importance of the C-C for Canada*	***Gatekeeper's importance index for Canada***
Calgary	**39.57%**	9.74%	**0.04334**	0.58%	**0.00256**
Saskatoon	17.57%	**27.63%**	**0.02993**	1.21%	**0.00132**
Edmonton	15.35%	18.47%	0.00510	1.36%	0.00037
Quebec	34.55%	7.97%	0.00267	0.28%	0.00009
Toronto	29.45%	6.16%	0.00204	**2.35%**	**0.00078**
Ottawa	11.51%	**29.94%**	0.00148	**2.56%**	0.00013
Vancouver	39.61%	3.01%	0.00097	0.34%	0.00011
Winnipeg	43.33%	6.49%	0.00046	0.14%	0.00001
Montreal	28.30%	5.01%	0.00030	1.05%	0.00006
Kingston	51.91%	4.95%	0.00029	0.14%	0.00001
Halifax	58.33%	15.74%	0.00004	0.16%	0.00000
Sherbrooke	100.00%	12.64%	0.00003	0.10%	0.00000
Average	**28.36%**	**10.52%**	**0.00522**	**1.47%**	**0.00050**

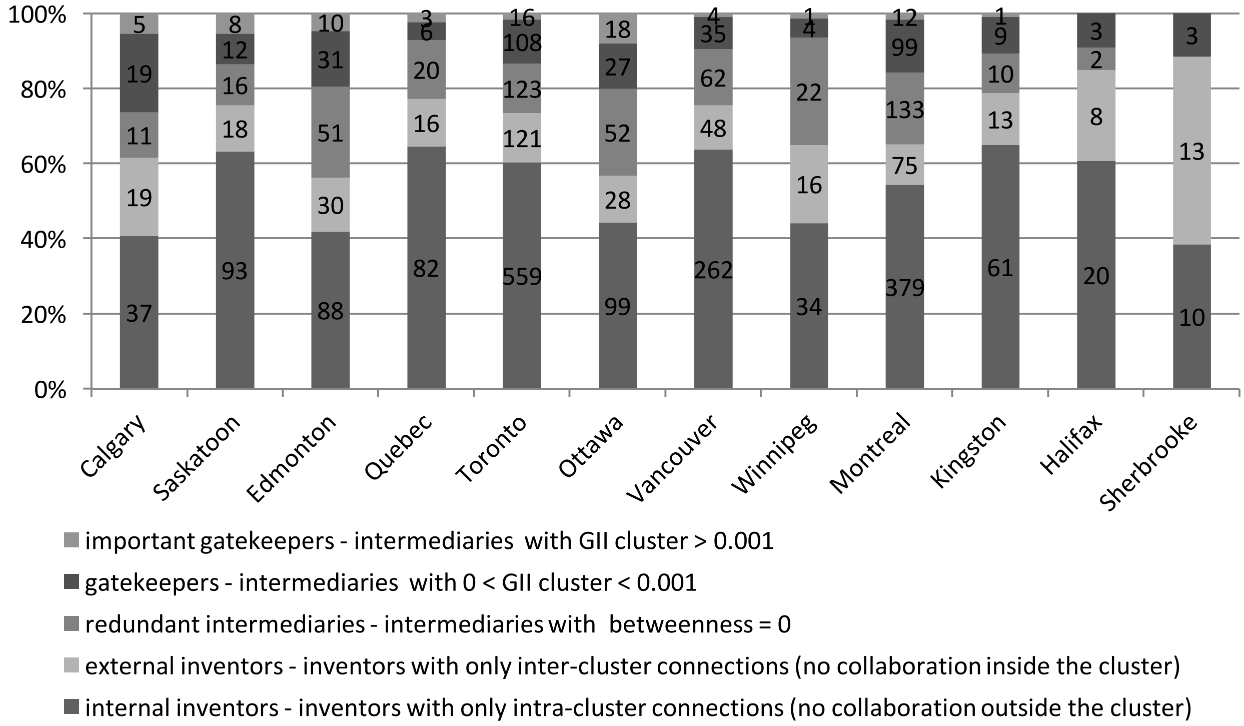


Figure 4: Numbers and relative proportions of inventors in the clusters categorized according to their importance as intermediaries measured by the Gatekeeper's Importance Index for the cluster