A Network Perspective of Nanotechnology Innovation: A Comparison of Quebec, Canada and the United States

Afshin Moazami

A Thesis

in

Concordia Institute for Information Systems Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Quality Systems Engineering) at

Concordia University

Montreal, Quebec, Canada

August 2012

© Afshin Moazami, 2012

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Afshin Moazami

Entitled: A Network Perspective of Nanotechnology Innovation: A Comparison of Quebec, Canada and the United States

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Quality Systems Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Chun Wang Chair

Dr. Zhigang Tian Examiner

Dr. Masoumeh Kazemi Zanjani Examiner

Dr. Andrea Schiffauerova Supervisor

Approved by ______________________________
Chair of Department or Graduate Program Director

____
Dean of Faculty

Date September 7, 2012

ABSTRACT

A Network Perspective of Nanotechnology Innovation: A Comparison of Quebec, Canada and the United States

Afshin Moazami

Given the novelty of nanotechnology science, its invaluable applications in almost all technological fields and its anticipated future effect on different aspects of economy and life, there is a need to study how the nanotechnology knowledge is produced. The main aim of this thesis is to compare the extent, the structure and the characteristics of knowledge transmission through innovation networks of collaborating nanotechnology researchers in Quebec, Canada and the United States, with a special focus on the interaction between academic and industrial researchers. We extracted the data from online databases of patents and articles, constructed the networks and applied the methods of social network analysis to compare collaborative patterns. The results have shed some light on various aspects of the knowledge networks. It was shown that the American nanotechnology network is more centralized; the researchers have more collaborators working in bigger research teams and engage themselves more frequently in university-industry partnerships. The Canadian network, on the other hand, relies more on purely academic research and is better interconnected internationally. Quebec's collaboration pattern is characterized by partnerships which are abundant but they take place within an increasingly more closed circle of Quebec academia. Furthermore, in all the regions it was found that academicians collaborating strictly within academia (and industrial researchers collaborating within industry) occupy more clustered network positions, whereas the researchers who collaborate across the university-industry boundary are more central, and thus critical for the effective knowledge transfer through the network. Based on our findings, several policy implications were derived for both Quebec and Canada.

ACKNOWLEDGMENT

Foremost, I would like to express my sincere gratitude and appreciation to my supervisor, Prof. Andrea Schiffauerova, for her support, understanding and unlimited generosity. I am grateful to her for giving me the opportunity to become a member of her research group. This thesis would not have been possible without her motivation and patient guidance. I am also thankful to Prof. Catherine Beaudry for her very valuable suggestions on the direction of my thesis.

I consider it an honor to have many friends who provided me a supportive and enjoyable environment in the past two years. I owe my deepest gratitude to Saba M. Heravi for her endless patience, forgiveness and kindness. Thanks to Ashkan Ebadi, and Elnaz (Tanaz) Kanani not only for being reliable co-workers, but also for being warm hearted friends. Special thanks to Parisa Mirshams, for encouraging me along the way. Credit goes to Babak Damavandi who never double checks his schedule for helping me. I am also indebted to my friends, Ehsan Khosroshahi Asl, Athena Kardehi Moghaddam, Mahdi Milani Fard, Ghazal Pasha, Javad (Arash) Mostofi, and Mohammad Reza Farhadi for their support, comments and feedback.

I would like to convey my thanks to current and former members of our research group in Concordia Institute for Information Systems Engineering for their contribution in data collection. I also thank Leila Tahmooresnejad who I frequently benefited from her knowledge.

Last, but by no means least, I owe my deepest gratitude to my beloved family, whose boundless love and continuous encouragement was unremitting source of inspiration for this work. No words can express how grateful I am for your love and support and how very much I love and appreciate you. This thesis is dedicated to them.

CONTENTS

List of Figures

List of Tables

1. INTRODUCTION

Nanoscale phenomena which had emerged over the recent years, are predicted to have an influential role in almost every aspect of the economy [1]. Nanoscale phenomena have two aspects; **nanoscience** which is referred to the underlying comprehension of atomic or molecular scale structure and operations, and **nanotechnology**, the process of manipulation and control of these phenomena to reach a concrete goal - building atom by atom [2] [3]. Although these two terms have different definition, there is no sharp distinction between them and both of them are often called key technology and science for the 21 century [1]. In literature, the term "nanotechnology" is used to avoid confusion and to be more appropriate [4].

There is a great amount of literature discussing the growth of this field [5] [6] [7] [8]. For example, it is mentioned that nanotechnology could generate major changes in the future as a "general-purpose technology[1]" [5]. Also, it is claimed that nanotechnology could bring high competitive advantage to most of the companies [6]. Furthermore, Shapira et al. [7] observed a shift from research to commercialization, which means that nanotechnology has found applications in industry. This is also confirmed by Freeman and Shulka [8] who report an increase in the amount of jobs in nanotechnology area.

The areas of applications of nanotechnology are very broad; they may range from medicine to biotechnology, aerospace, information technology and telecommunications [9]. As an example of wide application of nano-products we can mention "nanoparticles". In one hand, nanoparticles are used in scratch-resistant and light-resistant coating of windows and carbodies. On the other hand, light-sensitive nanoparticles are embedded in solar cells [10]. Moreover, Roco and Barinbridge [11] discuss the effects of nanotechnology in different areas from energy to ethics and public policy. For example, they argue that nanotechnology can decrease the input cost in

---

[1] This term is introduced by Bresnahan and Trajtenberg [114] to explain that such technologies would be a driver of modern economic dynamism.

some industries, and consequently increase the productivity. Also, it can increase the quality of human life by improving the renewable energy systems [11].

All these examples showing the future effect of nanotechnology on different aspects of economy and life suggest that nations need to start implementing the required infrastructure for adoption of the future changes in all related fields [12]. As a consequence, many countries such as United States, United Kingdom, France, Germany and Canada fuelled considerable resources and attention into the exploration of nanotechnology [2] [13]. In Canada, considerable amounts of funds have been dedicated to the nanotechnology through federal or provincial funding institutes [14]. However, this is not sufficient and Canadian nanotechnology research still struggles to find necessary funding [15]. Moreover, a comprehensive nanotechnology economic development strategy is lacking in Canada. As a result, Canada scores only as 13th in the creation of nanotechnology articles in the world [16]. The United States, on the other hand, have been very successful in terms of the nanotechnology innovation creation, which is the main reason why we suggest that Canada should look for inspiration in the US nanotechnology innovation system.

Given the importance of the nanotechnology innovation for the future development of Canada and so far not very developed Canadian nanotechnology innovation system, there is a need to study how the nanotechnology innovation is in fact created. We propose that the comparative analysis among the Canadian and the US systems in terms of the study of the knowledge flows among various knowledge producers at universities, research institutes and companies can help us to shed some light on the relation between the national innovation systems and knowledge diffusion within the systems. The main aim of this thesis is to compare the extent, the structure and the rate of knowledge transmission among nanotechnology researchers in Quebec, Canada and United States. We have included also Quebec in this analysis, because we, as Quebec residents, have a special interest in the local production of nanotechnology knowledge and innovation and would like to see whether it differs from the rest of Canada. To achieve our goal we have built the network of scientists and inventors and searched for the relations between the network architecture, the network position of the individuals and knowledge diffusion in the mentioned countries.

The remainder of the thesis is organized as follows. The following chapter presents a brief literature review in the field of national innovation systems and complex networks. In Chapter 3, the research objectives and our hypotheses are presented in details. Chapter 4 describes the data and methodology used for the analysis. We will see the methodology steps from nanotechnology keywords to nanotechnology innovation networks. Chapter 5 reports the results of our analysis in four sections; (1) network fundamentals, (2) regional characteristics of the networks, (3) academic and non-academic collaborations, and (4) regional comparison of academic and non-academic collaborations. Chapter 6 concludes, and Chapter 7 suggests some avenues for future research.

2. LITERATURE REVIEW

Literature review starts with the introduction of the concept of national innovation system. After a comparison of the Canadian and the US national innovation system, we will discuss the state of the art of nanotechnology based on the scientific publications.

In the second part, after introducing the concept of complex networks analysis, we will overview the research that deals with the co-authorship networks.

2.1 National Innovation System

2.1.1 Introduction

The term “National Innovation System”, defined as a conceptual framework in the science, technology and innovation studies, has been introduced in the late 1980s [17]. In the very first definitions of this term, the framework involved as a network of institutions that generate new technologies [18]. Later, some restrictions were added in order to narrow down its meaning and to make it more specific [19] [20] [21]. For example, Lundvall [22] suggested that these institutions should be either located within or rooted inside the borders of a nation state. Or, Patel and Pavitt [20] determined that the rate and direction of technological learning as a part of national innovation system definition.

According to Organization for Economic Cooperation and Development (OECD) [23], there are three factors that increase the importance on national innovation system in the technology field; economic importance of knowledge; increase in applying system approaches; and increase in the number of knowledge generation institutions [23].

National innovation systems may differ from one country to another from many aspects. Size of the country and its level of development, natural resources, governmental intervention and national financial system can make these differences. For example, in some countries, as for example in the US, the defense is a important part for the government and for the economy, while in non-defense oriented countries like Canada, telecommunication, equipment and energy have more priority. Also, rules and regulations can make differences. For instance, in European countries, an inventor cannot publish research revealing the methods behind innovations before their patenting, while it is allowed in the US [24].

The institutions that generate knowledge are categorized in three main sectors; universities, government labs and public or private industrial corporations. University units focus on fundamental research, while governmental labs and industrial sectors are mostly involved in applied research. Hence, universities mostly generate publications and patents, and they also train skilled personnel. The output of the government labs are prototypes, pilot plans and algorithms in addition to publications and patents. Finally, industrial sectors provide new products and processes or improve the existing parts [25].

2.1.2 The National Innovation System in Canada

Knowledge generation in Canada is performed by many institutions such as universities, companies, non-profit organizations and government laboratories. Learning processes and interactions among these institutes, which form the innovation system, can be either inter-industry or intra-industry [25].

According to the definition of national innovation system, the domestic interactions and processes are more important than international ones. However, in small countries like Canada, Sweden or Switzerland, the international linkages play much more important role compared to the larger countries like the US and Japan [26]. For example in case of Canada, integrating and applying the policies of the US and Western Europe has a significant role in structure of national innovation system [25]. Canada obtains more than 50 percent of its acquired technology from abroad [23].

Mcfetridge [27] studied the impact of the US on Canadian economy. He mentioned that even though a only small portion of the Canadian labor market is received from the US, the huge amount of imports and exports in Canada allows the US to influence Canadian companies and market. Also, agreements like NAFTA (The North America Free Trade Agreement) and FTA (Free Trade Agreement) between the US and Canada help to increase the collaboration between the Canadian and American companies.

RR&D in Canada is performed mostly by large Multi National Companies (MNCs) rather than small firms. The international partners of these companies, collaborating with them on innovative activities are most commonly found in the US, and then in some Western European countries, in Japan and in Australia [27].

There are two types of major categories in Canada's national innovation system: traditional and emerging innovation systems. The traditional innovation system usually involves domestic products and processes such as metallurgy, agriculture, energy and forestry, while the emerging innovation system includes aerospace, IT and telecommunications. There is a basic difference between these two categories. Traditional study fields and innovation system are highly dependent on natural resources, while the modern one mostly depends on human resources [25].

There are two important characteristics mentioned in the literature that differentiate Canada from other countries from the economy perspectives: first, strong government intervention in private national innovative activities [25]; and second, the integration with the much bigger US economy [27].

As mentioned before, there are three basic types of institutions that are involved in Canadian research and innovative activities: universities, public labs and firms performing industrial R&D. These three have different objectives; for example, commercial profit of cooperative research is the industrial firms' priority, while economic growth is the goal of government labs. However, to reach these goals, all of them need interaction, cooperation and technology transfer. Government encourages technological transfer because of higher chance of knowledge diffusion and technology production. The reason that research institutes and universities pursue research cooperation is the access to more resources and complementary knowledge. For industrial firm, on the other, the main motivation for research cooperation with other institutions is gaining new knowledge in order to increase commercial potential of its activities [25].

Stanley [28] defined several requirements for improving the efficacy of the Canada’s innovation systems. For example, in order to set up a system to promote the innovation, Canada needs to reorganize its science and technology funding in a federal level. Also, it is necessary that a federal innovation organization is founded to integrate the innovation production process along the country [28].

According to Stanley [28], there is a lag in adoption of new technology which affects the economy of Canada. Although, there are some exceptions, Canada lacks management capacity to lead companies on a global level. Comparing to developed countries, Stanley says that technology diffusion, both from abroad to Canada and within the Canadian innovation system, occurs slowly. To solve this problem, two solutions are proposed in Stanley’s article: to fortify the infrastructures in domestic market; and to maintain the open doors to foreign technologies and entrepreneurs [28].

2.1.2.1 Quebec Policy on Science and Innovation

A series of activities have been implemented by the Quebec Policy on Science and Innovation (QPSI) in order to improve the competitive advantages for Quebec’s economy. The followings are some examples of these activities [29]:

- Allocating of grants to start-up firms which have emerged from universities in order to obtain commercial benefits from academic research
- Performing action plans to manage the intellectual properties in the universities
- Improving the technological cooperation by supporting centers that focus on technology and knowledge transfer
- Supporting the regional innovation system to achieve a global competitive advantage

Even though Quebec's innovation system is dependent on the national research system, QPSI has its own activities, financial resources and infrastructures [29]. For instance:

- Reconstituting academic and research funding
- Funding infrastructures in research and development
- Developing high-tech industries like telecommunication and aerospace through investments in infrastructures and R&D
- Developing international networks to increase the international impact of Quebec research
- Increasing the responsibility of scientists in developing the society’s values
- Raising citizens' awareness to facilitate further innovations
- Training the qualified human resources

2.1.2.2 Nanotechnology Policies in Canada

It is estimated that the worth of nano products will be $1 trillion in 2015 and its great impact on economy and its rapid growth is a concern for many countries; Canada in not an exception. Nanotechnology is called a “platform” technology like internet and electricity. In other words, Nanotechnology is predicted to direct the future technological change in the society [30].

In Canada, public policy frameworks manage the risks, benefits and its ethical, environmental, economic, legal and social impact. These policies consider the resources and priorities and develop a database of all desired information and regulation to ensure that all significant aspects of nanotechnology are recognized [30] [31].

Nanotechnology, biotechnology, and information and communication technology (ICT) are three technologies that government of Canada considered in their comprehensive Science and Technology Strategy in 2007. Also, Canada participates in the International Organization for Standardization (ISO) to develop nanotechnology standards. Furthermore, a large portion of activities on nanotechnology regulation to be addressed in cooperation with OECD countries are assigned to Canada. [32]

According to Kuroiwa [33], about half of the institutes of the National Research Council (NRC) conduct R&D in nanotechnology. Also, the number of nano-related companies in Canada is estimated from 50 to 200 depending on the definition used for the term “nanotechnology” [33].

In order to facilitate the commercialization of nanotechnology and to allocate funding to R&D in this field, the National Institute for Nanotechnology (NINT) was founded in 2001 through a partnership between the University of Alberta, the government of Alberta and the NRC.

There are several challenges for Canadian policies dealing with nanotechnology, such as lack of a nanotechnology economic development strategy and a lack of necessary funding compared to the US, Japan and Europe. According to government in 2002 [15], the US invests nearly six times more per capita in nanotechnology than Canada. Nanotechnology research in Canada struggles to find necessary funding and there is a great time lag between finding the suitable funding programs, applying for the funding, receiving the results of decision process, release of funding and, finally ordering and receiving the desired equipment [15]. Moreover, another challenge for Canadian nanotechnology is the lack of large companies involved in nanotechnology [15], which is usual in other countries with successful nanotechnology innovation, e.g. the United States.

In the next section we will look at the national system of innovation in the US in order to highlight the differences between the two systems.

2.1.3 The National Innovation System in the US

In the US, similarly as Canada, different institutions are involved in nanotechnology research including companies, government agencies, government research laboratories, universities, and non-profit organizations [34].

Between 1991 and 2000, the US accounted for 43 percent of the total R&D among the OECD countries[2] [35]. The R&D expenditures are divided into three categories:

- Development, i.e. "the production of useful materials, devices, systems, or methods, including the design and development of prototypes and processes". Development accounts for 60-65% of total R&D over the last two decades[3] [35].
- Applied research, i.e. a specific need or commercial objective in products, services or processes. It is claimed that applied research accounts for 22% of R&D [35].
- Basic research, i.e. Fundamental knowledge without considering a specific application. This category has the smallest share in R&D (18%) [35].

The US has the largest innovation system among all the OECD countries. The roles of innovation performers - government, industry and universities - have changed over the past 70 years. The significant role new firms play in the commercialization of new technologies such as biotechnology, computer software and hardware, microelectronics and robotics over the past four decades is one of the important characteristics of the US innovation system. Moreover, small firms are very active in the commercialization of new technologies, which is also an important feature of the US innovation system [36].

In addition to institutions performing R&D, the US national innovation system involves a wide range of the policies, such as antitrust policy, intellectual property rights, and regulatory policy [37]. Simons and Walls [34] have mentioned a series of public policies that affect technology development in the US, such as the training of scientists and engineers, and technology adoption, which have a significant impact on the success of the national innovation system. Another policy, known as "Bayh-Dole Act" of 1980, is considered to be one of the most important policies, which changed the face of the

---

[2] Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, and United States.

[3] Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, and United States.

academic innovation in the US. Bayh-Dole Act allows the universities, non-profit organizations and small businesses to patent and later to commercialize the results of their research, even if the research has been funded federally [34]. In case of universities, this policy increased the share of patenting from less than 0.3% in 1963 to nearly 4% by 1999 [37].

2.1.4 Nanotechnology Publications and Patents

Publications are known as outcome of a scientific research indicating the new findings. Since scientific research is the basis of nanotechnology, publications can indicate its progress. Number of publications is considered to be a good indicator for the growth rate of nanotechnology [38]. According to Miyazaki and Islam [4], universities have the largest portion of publications in nanotechnology among all the institutes (70.45% of nanotech related articles.) Public research institutes with 22.22% of all publications follow the universities, while private sector with only 7.3% has the smallest publication share. Also Chen and Roco [39] observed that the most productive institutes were universities and national research centers rather than private companies (all top 20 institutes from 1976 to 2004.) These numbers are not surprising, because nanotechnology is in its emerging phase and most of the publications are in basic research, and due to the high-risk aspect of new technology the private sector is reluctant to be involved in the basic research [4].

Number of publications only shows the quantity of the knowledge production, while an indicator for the quality of this knowledge is needed as well [40]. Number of citations of the published article is a universally accepted indicator showing the quality of publications. Using both the number of publications and the citation count, Youtie et al. [41] compared the nano-related publication of different countries since 1990 through 2006, and identified the leading countries. Based on their result, the US and European Union 27 (EU27) have largest number of nanotechnology publications, while some Asian countries like China and Asian Tigers (South Korea, Singapore and Taiwan) greatly increased the publications recently with a higher rate than others [41]. Also, Hullman

[38] categorized the nano-related articles based on the countries in two time intervals (1992-1995 and 1998-2001). Her results suggests that the US and Canada share in publication is decreased in the second interval, while Asian countries are get more dynamic. Hullman compared different countries based on the publication citations. Using "cite per paper", the relative impact is calculated. Results indicate that Switzerland and Netherlands are leading, and the US and Canada are following them [38].

As reported by Roco [42], since 2000 though 2008, there is a 23% annual growth in nanotechnology publications. Also, Chen and Roco [39] observed that 213,847 articles are published in 4,175 journals between 1976 and 2004 analyzing the Thomson Science Citation Index (SCI). Among them the US with 61,068 articles has the largest amount of nano-related publications, followed by Japan (24,985), Germany (21,334), and China (20,389) [39].

Also, an increase in the number of nanotechnology patents is reported [38]. Similar to publications, the US has the first rate in the number of patents in nanotechnology, followed by China, Japan, South Korea, and Canada [43].

2.2 Networks

2.2.1 Complex Networks

In general, a network is a set of items, called vertices[4] that are connected to each other with some edges[5] [44]. Different examples of networks can be found in the nature; people as units of a network of different kinds of social relationships; web pages as vertices of World Wide Web that are connected through the hyperlinks; the network of business relations between companies; the power network of a country; neural networks; citation network between scientific papers, etc. [45] [46].

---

[4] Also called "node" in computer science, "site" in physics, or "actor" in sociology [44]

[5] Also called "link" in computer science, "bond" in physics, or "tie" in sociology [44]

Historically, the study of networks is based on Graph Theory; a branch of discrete mathematics[6] [44]. In mathematics, a network consists of a graph of N vertices connected by a set of M edges and additional information on the vertices or the edges of the graph [47]. For each vertex i, d_i is defined as the degree, which is the number of edges connected to that vertex. This graph is represented by an adjacency matrix A in which entry $a_{ij} = 1$ if there is an edge between vertex i and vertex j [48].

Complex networks are defined as networks with irregular structure and thousands or millions of vertices which are evolving dynamically and in a complex way. In complex networks, which became the focus of attention in the last decade, the main analyses are on the properties of dynamic units of the network [45].

Recently, applications of complex networks are found in many areas such as sociology (e.g. [49] [50]), economics (e.g. [51] [52]), biology (e.g. [53] [54]) and scientometrics (e.g. [55] [56]) The major reason for the popularity of complex networks can be the availability of large datasets [57]. Thanks to high performance computers and communication networks, gathering and analyzing data on a much larger scale than previous is nowadays possible [44].

According to Newman [44] small networks and complex networks differ in two aspects. First, in small networks, the role of each vertex or edge is studied. For example, it is questioned whether the removal of a specific vertex can change the network's connectivity or not. While in complex network with millions of vertices, a single vertex (or edge) is not a concern. In other words, instead of studying the effect of removal of one vertex, for example, the removal of a percentage of the vertices is examined. As a consequence, instead of exact questions, statistical questions are addressed in the problems of complex networks. The second difference is that small networks with tens or hundreds vertices can be drawn with actual vertices and edges, but for complex

---

[6] The Königsberg bridge problem is known as the birth of the Graph Theory. This problem, which is solved by Swiss mathematician Leonhard Euler in 1736, consists in finding a round trip that traversed each of the seven bridges of the city of Königsberg in Prussia exactly once. For more details about the graph theory see [61]

networks statistical approaches are needed to answer questions like "How can I tell what this network looks like, when I cannot actually look at it?" [44].

Complex networks analysis has three steps. (1) Developing **models** to understand the topology of the networks. (2) Finding statistical **properties** to characterize the structure and behavior of networks. (3) **Predict** the behavior of networked system based on the model structural characteristics of that network [44].

In the following section, we will discuss the network models and their statistical properties that we used to predict the behavior of our networked system in this research.

2.2.1.1 Complex Networks Models

Several models are proposed for studying the topological properties of the complex networks. Some of them like random graph, small-world model, and scale-free networks are general models that became subject of great interest [57]. Apart from the general models, there are also models applied to specific networks, but these will not be covered here.

The **random graph model**, proposed by Erdős and Rényi in 1959 is one of the most basic models of complex networks. According to this model, we start with large number of isolated vertices (n), and randomly add edges between them until we have an average of one edge per vertex. Another representation of this model, known as Erdős and Rényi model (ER model), gives a probability of p to presence of each edge. It is proven that the histogram of this model follows a Binomial distribution and for sufficiently large n and small p, it follows Poisson distribution [57].

The **small-world model** was developed by Watts and Strogatz based on "small-world" property and high clustering coefficient. The "small world" property originated from the Milgram's experiment in 1967 who found that two random US citizens were connected to each other on average by 6 acquaintances [58]. In this experiment, a number of letters to a random selection of people in Nebraska were distributed. Each letter has instructions to send them to the addressee in Boston by passing them from person to person. Although, it was expected that each letter will need to go through around 100 persons in order to reach the addressee, Milgram found that it had only taken an average of 5.5 persons to get from Nebraska to Boston. This number rounded up to 6 and became known as “six degrees of separation” [46] [58].

Clustering coefficient C is defined as the probability that two vertices will be connected, given that each is also connected to a mutual friend. Clustering coefficient is obtained by dividing the number of actual edges between one’s friends (E_i) by maximum number of edges that they could have:

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

where k_i is the number of friends of vertex i [59].

Combining these two ideas, Watts and Storgatz [59] developed the small-world model. To construct a small-world network, we start from a circle of vertices where each vertex is connected to the nearest neighbors, which results a high value for clustering coefficient. Then a few long-range edges connecting randomly selected vertices, drastically shortens the average separation between all vertices [46].

The random model of Erdős and Rényi rest on two simple assumptions; first, all the vertices available from the beginning (vertices are not created or destroyed); second, all the vertices have equivalent chance to get edges. Barabási and Albert developed **Scale-free Networks** by changing these two assumptions to two new assumptions; growth and preferential attachment. In their model, at each step network *grows* with the addition of new vertices during its construction. Also, the most connected vertices have greater chance to get edges to these new vertices. This property called *preferential attachment*, also known as “rich get richer” paradigm [57].

In Figure 1 from [57] the difference between these three models is illustrated. The degree distribution in two first models are almost the same (with a peak), while the third model suggest that the degree distribution follows power law.

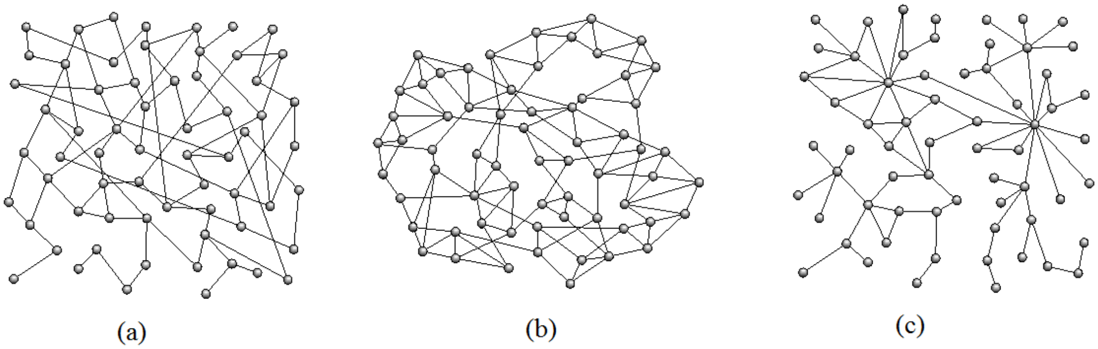


Figure 1 - Complex networks models (a) an example of Random Network by Erdős and Rényi, (b) an example of Small World Model By Watts and Strogatz, and (c) an example of scale free network by Barabási and Albert. Figures are taken from [57].

2.2.1.2 Network Structural Properties

The most important structural properties of networks can be categorized into five groups; vertex centrality, the small-world effect, transitivity or clustering, degree distributions, and fragmentation.

Vertex Centrality

Centrality measures are among the most popular measurements that show the importance of a vertex or edge in the network from different aspects [60]. These measures are typically categorized based on the type of property that they are related to, such as shortest path, degree and distance. In the following section, the measures that will be used in this study will be briefly explained.

- **Betweenness Centrality**

Betweenness centrality is a measure based on the shortest path. In graph G, for a pair of vertices like s and t, shortest path from s to t is defined as an order of adjacent vertices with minimum number of edges that connects s to t [61].

Betweenness centrality is an indicator that is defined for each vertex like v as sum of the fraction of all-pairs shortest paths that pass through v, over all-pair shortest paths [62]. In other words, in order to calculate the betweenness centrality for a vertex like v, we need to consider all the pairs like s and t in the network, and sum up the fraction of the number of shortest paths from s to t that pass through v over the total number of shortest paths from s to t. Mathematically, it is defined as

$$\mathrm{C_B(v)} = \sum\nolimits_{\mathrm{s \neq v \in V}} \sum\nolimits_{\mathrm{t \neq v \in V}} \frac{\sigma_{\mathrm{st}}(\mathrm{v})}{\sigma_{\mathrm{st}}}$$

where $\sigma_{\mathrm{st}}(\mathrm{v})$ denotes the number of shortest paths from s to t that pass through v, and σ_{st} is the total number of shortest paths from s to t . This measure was introduced by Freeman in 1977 [63] and Anthonisse in 1971 [64].

Although faster algorithms based on breath-first search (BFS) are introduced [62][7], the calculation of betweenness centrality in large networks is still very time consuming.

- **Degree Centrality**

As mentioned in 2.2.1, the degree is the number of edges connected to one vertex. Degree centrality of a vertex is simply defined as the degree of the vertex. The average degree centrality is used to compare the networks or sub-networks in terms of connectivity [49].

- **Closeness Centrality**

Closeness centrality is based on the idea that more central a vertex is, more quickly it can interact with other vertices. In other words, a vertex is more central if it is on average closer to other vertices. This indicator is defined for each vertex based on its

---

[7] It takes O(VE) time.

distance to other vertices. In graph G, the distance between vertices u and v is defined as the minimum number of edges that needs to be traversed from u to reach v. In more formal words, the distance between u and v is defined as the length of the shortest path between these two vertices, and if there is no path between these two vertices the distance is assumed to be infinity or undefined [61]. As a consequence, closeness centrality can be measured only in connected networks, i.e., in the network where all the vertices are directly or indirectly connected.

Closeness centrality for vertex v is defined as inverse of summation of distances from v to other vertices in the graph. It is formally defined as

$$C_C(u)^{-1} = \sum_{i=1}^{n} d(v_i, u)$$

where $d(v, u)$ is the distance of v and u. In other words, a vertex is closer if it is more accessible from all other vertices [60]. An example application is shopping mall locating problem in which we want to locate a shopping mall that is close to all clients in total [65]. This measure was introduced by Sabidussi in 1966 [66].

The Small-world Effect

We mentioned the Milgram's experiment in 2.2.1.1 which proved the "six degree of separation" in social network of the US citizens. Formally, the small world effect in an undirected graph defined as the mean of shortest distance between pair in the network (l):

$$l = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}$$

where d_{ij} is the distance between i and j and n is the number of vertices in the network[8].

Many experiments calculate l for different complex networks. The degree of separation for network of film actors with about 450K actors and 25M edges (defined by the collaborations in films) was calculated as 3.48. For the network of protein interactions the l is 2.12. The small-world effect in the World Wide Web was calculated between 11 and 17, depending on which fractions of the websites are included in the dataset [44].

Transitivity or Clustering

The concept of transitivity or clustering introduced already in 2.2.1.1 concerns the presence of an elevated number of triangles in the network. Triangles in a graph are represented by three vertices that are all connected. In other words, we call a network more transitive or clustered, if it is more probable that a friend of a vertex's friend is its own friend as well [44].

Clustering coefficient as an indicator of transitivity has different definitions, but the one that proposed by Watts and Storgatz [59] is widely used. In this definition, a local value C_i shows the clustering coefficient for vertex i which is the ratio of "number of triangles connected to vertex i" to "number of triples centered on vertex i". The clustering coefficient of the network is the average of C_is.

Generally in real-world networks, the values of clustering coefficient tend to be considerably high comparing to random networks [44].

Degree Distributions

Another important property that is defined in complex networks is degree distribution. If we define the fraction of vertices in a network with degree of k as p_k, the histogram

---

[8] In this definition the distance from each vertex to itself is included. Some definitions do not include this [61]. However, it does not affect the result.

of p_k shows the degree distribution of the network. The cumulative degree distribution is defined as follows:

$$P_k = \sum_{k'=k} p_{k'}$$

which shows the probability that the degree is greater than or equal to k [44].

Many real-world complex networks follow power laws: $p_k \sim k^{-\alpha}$ where α is a constant exponent. Consequently, the cumulative degree distribution follows the power law with exponent of $\alpha - 1$. Figure 2 from [44] illustrates the degree distribution of some complex networks.

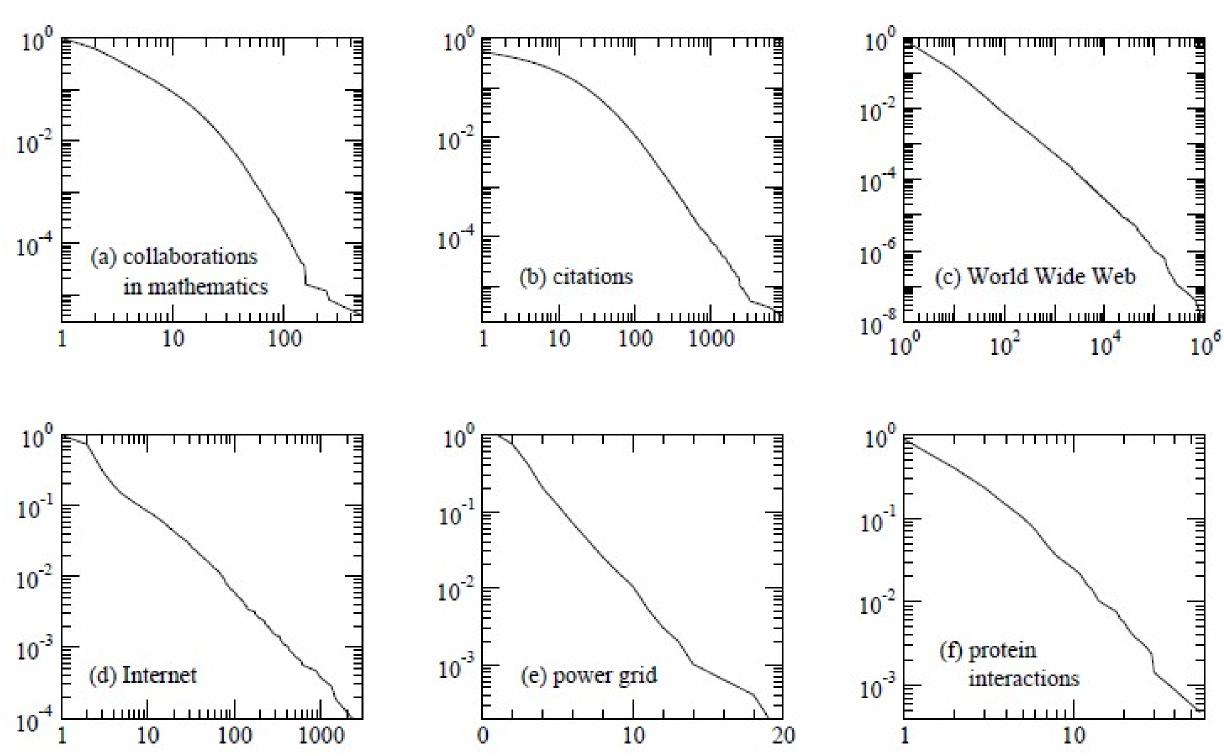


Figure 2 - Cumulative degree distributions (P_k). Horizontal axis is vertex degree and vertical axis is cumulative probability. (a) The co-authorship network of mathematics, (b) citation network of articles, (c) WWW, (d) Internet at autonomous systems level, (e) the western US power grid, (f) the proteins of metabolism [44]

Fragmentation

Fragmentation shows how vertices in the network are connected to each other. There are several indicators that show the level of fragmentation in the network, such as size of the largest component, average size of components, and number of isolated vertices.

Component in a network defined as a maximal subset of vertices that are connected directly or indirectly by edges [61]. Also, isolated vertices or components of size 1 refer to the vertices with no edges. A network with fewer vertices in the largest component, smaller value of average size of components, and more isolated vertices is more fragmented [67].

After introducing the complex networks, in the next section, we are going to focus on a common type of complex networks, called collaboration network that is used in this thesis.

2.2.2 Collaboration Networks

One of the applications where the concept of the complex networks is used is collaboration networks, which are also is the focus of this research. In these networks the actors (vertices) are individuals and two actors are connected to each other if they collaborate in a product [68]. The main aim of this thesis is to study the structure of the innovation networks, in which we will consider two types of linkages; first, collaborative links defined by the co-authorships of scientific articles among scientists working on joint research projects, and second, links defined by co-inventorship of patents among inventors who made the patentable discovery.

2.2.2.1 Co-authorship Network of Scientists

The idea of the co-authorship networks comes from the Erdős number introduced by Goffman in 1969 [69]. For each scientist he calculated the Erdős number, which is the length of the shortest path from that scientist to Paul Erdős[9]. This means that Erdős himself has number 0, and his co-authors have number 1 and so on [70].

Increasing the interest to the complex networks, the Erdős number has been elevated to more serious scientific subjects [71]. It attracted the attention of two types of scholars: those who study social networks and those who study the reasons and consequences of collaboration [72].

First group is interested in this topic because the co-authorship network of scientists is one the largest social networks and it represents a prototype of dynamic complex networks [56]. Newman [73], for instance, has constructed co-authorship network of scientists in many disciplines in a 5 years period of time (1995-1999). He measured basic measures for this network such as numbers of papers written by authors, numbers of authors per paper, numbers of collaborators that scientists have existence and size of a giant component of connected scientists, and degree of clustering in the networks. Later he extended his work to some advanced measures like distances between scientists through the network, and measures of centrality such as closeness and betweenness [74]. Also, Barabási et al. [56] studied the time evaluation of this type of complex networks using a dataset on publication on mathematics and neuroscience in a period of 8 years (1991-1998). According to them, the degree distribution of their collaboration networks follows the power law. They also observed the relation of some key quantities with time. For example, clustering coefficient decays with time, average degree increases (approximately linearly) [56].

On the other hand, scientists who study the collaboration focused on the phenomenon of co-authorship because publications are important indicator of collaboration in academia [72][10]. For example, Moody [75] studies a network of scientists from 1963 to

---

[9] "Paul Erdős (1913-1996) was a Hungarian mathematician who published more papers than any other mathematician in history working with hundreds of collaborators." (From Wikipedia: http://en.wikipedia.org/wiki/Paul_Erd%C5%91s)

[10] Co- authorship is an approximate partial indicator of collaboration and using this indicator needs some cautions. Because, many cases of collaborations are not necessarily results in a co-authored publication [115].

1999 to show that participation in the sociology collaboration network depends on research specialty. Also, Glänzel and Schubert [76] observed some collaboration benefits using the analysis of the collaboration network. According to them, co-authorship (and cooperation in general) appears to be ‘cost effective’ on the long run. Also they suggested that collaboration should be encouraged and supported because it is able to promote research activity, productivity, and increases the impact of the resulting work [76].

Also, Acedo and Barroso [72] studied the co-authorship network in management and organizational studies. They compared this discipline to other disciplines and showed that articles with two authors seem to have a greater impact. Moreover, there are some other studies of co-authorship networks that focus on different levels of the network such as individuals, cross nation level, and multi-national collaborations [76]. Another example is measurement of the impact of university-funded research and collaboration on scientific production of Canadian biotechnology academics by Beaudry and Clerk-lamalice [77].

2.2.2.2 Co-inventorship Network of Inventors

The network of patent co-inventorship has also similar structure – two inventors are connected if they collaborate in a patent as joint inventors. The idea of how to construct this network was presented by Breschi and Lissoni [78] [79] and later by Balconi et al. [80], and consists in linking Italian inventors using data on co-inventorship of patents registered at European Patent Office (EPO). They construct a bipartite graph of patents from 1978 to 1995 and their inventors, and based on that they generated the co-inventorship network of Italian inventors. Various measures of social proximity between cited and citing patents and many centrality measures of this innovation network were measured in these studies.

Other developments of co-inventorship networks, or in general innovation networks, are also proposed. For example, Cantner and Graf [81] established the network of innovators characterized by innovators’ technological fields, which means two innovators are linked whenever they patented in the same technological class. Moreover, Singh [82] inferred collaborative links among inventors using social proximity graph, which he also constructed from patent collaboration data.

Similarly as was discussed in the section on the article co-authorship networks, also here many researchers adopt the patent co-authorship as an appropriate social relationship between the patenting inventors to study the collaboration effects [16] [78] [79] [83] [84]. For example Beaudry and Schiffauerova [84] studied the impact of collaboration and co-inventorship network characteristics of Canadian nanotechnology inventors on the quality of their inventions. Also, Guan and Shi [85] focused on the network effect on the system and individual level of the innovative creativity.

Although this method is used widely, some researchers indicate its disadvantages. For instance, Fleming et al. [86] warned that links in co-inventorship networks differ significantly in their strength and information transfer capacity. Also, a considerable portion of the old relations remain viable even if the tie does not exist anymore.

There has been a lot of research performed on the subjects of national innovation systems and collaboration networks, though these topics have not been studied together. The main purpose of this thesis is thus to shed some light on the knowledge flows within innovation systems of two countries (Canada and the US) and relate the collaboration network structure to the characteristics of the national innovation system in each country. In the following section we propose several hypotheses to that effect.

3. RESEARCH OBJECTIVES AND HYPOTHESES

3.1 Hypotheses

3.1.1 Basic Characteristics of the Network Structure

There is an increasing academic interest in the phenomenon of collaboration among scientists. Like other collaboration networks, social network analysis approach is applied on this network to identify its characteristics. Newman [73] observed the small-world effect on collaboration networks and explained the properties of this type of networks using the small-world model of Watts and Strogatz [59]. Also, Barabási et al. [56] studied the evolution of these networks over the time, and claimed that the collaboration networks follow the scale-free model, known as Barabási-Albert Model [46].

- ***Hypothesis 1a:*** *Collaboration network of scientists who have joint article or patent in the field of nanotechnology can be described using the small-world model.*
- ***Hypothesis 1b:*** *The evolution of collaboration network of scientists follows the properties of scale-free networks, and can be modeled by Barabási-Albert Model.*

3.1.2 Regional Characteristics of the Network Structures

Canada became involved in nano science and technology later than the US and it still has not reached the similar level of development as the US. According to the Canadian Workshop on Multidisciplinary Research on Nanotechnology [15], Canada lacks a nanotechnology economic development strategy and lags in nanotechnology funding compared to the US. Nanotechnology research in Canada struggles to find necessary funding and there is a time lag to seek funding programs [25]. The US innovation system, on the other hand, is larger than that of Canada and of other OECD countries, it supplies significant financial resources for both basic and applied research, encourages university-industry collaborations and provides critical public policies on federal and state level.

For the comparison between the Canadian and the US nanotechnology networks we anticipate to find that the Canadian network is much smaller and much less developed (even if adjusted for size). American network is thus expected to exhibit more favorable properties for the knowledge transmission among the researchers.

- ***Hypothesis 2:*** *American nanotechnology network is bigger, more dense and compact and involves more collaboration than Canadian nanotechnology network.*

Canada has a small population dispersed over a large geographical area and its private sector is dominated by small-sized and medium-sized companies. As a consequence, research and development has to concentrate in geographical agglomerations and clusters in order to contribute to an efficient innovation system. In nanotechnology, most of the research in Canada is concentrated in the clusters of Toronto, Ottawa, Montreal and Vancouver [67]. The study on the geographical aspects of the collaboration pattern based on Canadian nanotechnology patenting [87] shows that most of the collaborations take place within the clusters or within the short geographical distance around them, or with the international partners, mainly from the US. Inter-cluster research partnerships seem to be much less interesting for Canadian inventors.

Therefore we expect to confirm through our network collaboration analysis that foreign participation and international linkages (mainly directed towards American researchers) are very important parts of the Canadian innovation system.

- ***Hypothesis 3:*** *The collaborations of Canadian researchers with their international (mainly American) counterparts form a significant part of the overall Canadian collaboration pattern.*

Even though Quebec is a part of the Canadian innovation system, we assume that it is in a slightly different position, mainly because of the language aspect. We expect that the

Quebec network will exhibit similar network properties as the Canadian one, but assume that due to the language issue the network will be more closed. Compared to the researchers from other provinces, the Quebec researchers are expected to collaborate more with each other, and to create more stable research relationships with many repetitive collaborations between the came collaborative pairs. The collaborations with Anglophone world are expected to be relatively less numerous.

- ***Hypothesis 4****: The collaboration of the Quebec-based researchers involves more internal research relationships within Quebec, which tend to be more repetitive, and less collaboration with the Anglophone researchers, compared to the collaboration pattern in the rest of Canada*

3.1.3 Collaboration between Academia and Industry

Even though it is often argued that universities frequently exchange information with the private companies and other organizations, the knowledge transfers from university-based open science to commercial science are quite inefficient. Dasgupta and David [88] described the differences between the social organization of the world of science (characterized by publication and supported by a priority-based reward system) and the world of technology (in which ideas are produced for economic objectives and encoded in patents). A consequence is a constant friction between these two distinct groups. Murray [89] has argued that only few key scientists publish across industry-academic boundaries and firms in fact rarely participate in science. Zucker *et al.* [90] confirm that especially among scientists it is commonly thought that the very best scientists are unlikely to be involved with the firms or to patent their discoveries.

Nevertheless, in the field of nanotechnology, in which the technology significantly relies on the science, we assume that there must be a much more collaboration between universities and companies. We propose that with our methodology we will be able to better track the missing interactions between the academic and industrial innovators and provide evidence of the significant collaborations across industry-academic boundaries.

- ***Hypothesis 5:*** *Academic nanotechnology researchers frequently carry out joint research activities in conjunction with both academic and non-academic organizations.*

Innovation networks of academic and non-academic researchers exhibit distinct collaboration patterns. Hence, we expect that the structural properties of the network nodes will depend on the affiliations of the subjects (academic or industrial scientists). Balconi *et al.* [80] observed that networks of inventors within industrial research are usually highly fragmented, while the academic networks constructed by Newman [73] were highly clustered. Newman [74] also argued that for most scientific authors the majority of the paths between them and other scientists in the network go through just one or two of their collaborators. This is in agreement with Balconi *et al.* [80] who found that academic inventors that enter the industrial research network are, on average, more central than non-academic inventors - they exchange information with more people, across more organizations, and therefore play a key role in connecting individuals and network components. Academics also have a tendency to work within larger teams and for a larger number of applicants than non-academic inventors [80].

Therefore we expect to find similar results: the networks of academic researchers will be highly clustered and very centralized, whereas the networks of industrial researchers will be more fragmented and less centralized. Moreover, we suppose that academicians will be found to work in larger teams, while industrial networks will be composed of smaller teams.

- ***Hypothesis 6:*** *Academic nanotechnology subnetworks are more clustered, more centralized and the nodes have higher number of direct ties than non-academic nanotechnology subnetworks,* It was already suggested [80] that the networks of industrial inventors are much more fragmented than networks of academic inventors. The fragmented non-academic networks are however bridged together by academic researchers who connect the industrial network components, and thus serve as connecting agents between the different organizations within the world of technology.

We expect that our analysis will confirm the hypotheses on the significant position of academic scientists in collaboration with industry. According to Breschi and Catalini [91], the authors of scientific articles who are at the same time also patent inventors have prominent positions in networks, compared to the positions of only scientific authors or only patent inventors. This shows that the science and industry collaboration have positive impacts on the scientific and collaboration networks. Hence, we expect that the collaboration of academic scientists with industry enables them to occupy more important positions within their networks.

- ***Hypothesis 7a:*** *Academic nanotechnology scientists, who co-author articles with industrial scientists, occupy more cliquish positions in the co-authorship network compared with academic scientists who do not collaborate with industrial scientists.*

- ***Hypothesis 7b:*** *Academic nanotechnology scientists, who co-author articles with industrial scientists, occupy more central positions in the co-authorship network compared with academic scientists who do not collaborate with industrial scientists.*

3.1.4 Regional Differences in the Collaboration between Academia and Industry

Since we aim to explore the university and industry scientists' collaboration in nanotechnology within different areas including Quebec, Canada and the US, the previous hypotheses related to the academic and non-academic subnetworks and collaborations (Hypotheses 4, 5 and 6) will be tested for Quebec, Canada and the US, and then comparisons of Quebec-Canada and Canada-US will be performed. Moreover, the following hypotheses will test the different conditions within the US national innovation system and the Canadian one. We propose that there are fundamental differences related to the collaboration between academia and industry, which should be reflected in the network structures:

Universities in the US are encouraged to collaborate more with industries. This collaboration facilitates the knowledge flow between these two sectors and streams industry funding toward university research [25]. Companies, on the other hand, can exploit universities' knowledge as they are important sources for basic research and technologies [34]. The collaboration with universities helps companies to reduce their R&D costs and to explore new areas of scientific research. According to Hill [35], the American private industry has changed the way it conducts R&D since some firms have decided to externalize part of their R&D to universities and research institutes.

This system is still not as developed in Canada. Universities appear to be effective in basic research as they produce a reasonable number of research papers compared to other OECD countries [32]. However, Canadian universities contribute less to domestic industrial research and the funding of business sectors in university R&D is also smaller than in the US. The majority of university funding is provided through government grants and contracts, while the industry accounts for only a very small part. The majority of industry research is funded by corporations themselves [25].

Therefore we expect that the collaboration pattern between the academic and non-academic researchers will be quite distinct in both countries, showing more developed academia-industry relationship in the US. We propose that this will be also reflected in the network positions of the academic researchers in both countries.

- ***Hypothesis 8a:*** *American academic nanotechnology scientists collaborate more with non-academic scientists, compared to their Canadian counterparts.*

- ***Hypothesis 8b:*** *American academic scientists occupy more central network positions than the Canadian ones.*

- ***Hypothesis 8c:*** *American academic scientists occupy more cliquish network positions than the Canadian ones.*

Large companies play a critical role in developing the market share of new products, but Canadian large companies are not very much involved in nanotechnology. Canadian industrial research is more concentrated in the small fraction of firms, which causes innovative activities to be more specialized. The nanotechnology research is thus performed only within a few firms [87]. As there is much greater involvement of American large companies in nanotechnology we expect the American industrial nanotechnology network to be much greater, dense and more developed.

- ***Hypothesis 9a:*** *American non-academic nanotechnology network accounts for a greater proportion of the researchers than the Canadian one.*

- ***Hypothesis 9b:*** *American non-academic nanotechnology network is more centralized and clustered than the Canadian one.*

3.2 Objectives

Objective 1: Create a methodology and tools for the data extraction and database building.

- Create an automated tool for the extraction of the data from the Internet-based databases of scientific articles based on keywords (which will be used for collecting data in future research by other students)
- Build a database of the world nanotechnology scientific publications (which will be used for further research by many other students)

Objective 2: Investigate the collaboration characteristics of scientists (inventors) through co-authorships (co-inventorship) networks in Quebec, Canada and the US

- Determine the collaboration characteristics and the network positions of scientists (inventors) through their co-authorship (co-invention) linkages in Quebec, Canada and the US.
- Compare the collaboration characteristics and the network positions of scientists (inventors) within the co-authorship and co-inventorship networks of Quebec, Canada and the US.

Objective 3: Investigate the collaboration characteristics of academic and non-academic scientists through co-authorships networks in Quebec, Canada and the US

- Determine the collaboration characteristics and the network positions of academic and non-academic scientists through their co-authorship linkages in Quebec, Canada and the US.
- Compare the collaboration characteristics and the network positions of academic and non-academic scientists within the co-authorship networks of Quebec, Canada and the US.

4. DATA AND METHODOLOGY

4.1 Data Extraction

Since we are interested in both academic and industrial aspects of innovation production we need to characterize both co-authorship network of scientists and co-inventorship network of inventors. For this purpose the information related to both nano-related articles and their authors, and nano-related patents and their inventors needed to be gathered. We encountered numerous challenges during the data gathering process, which are in detail described in this chapter.

4.1.1 Articles Database

Gathering nanotechnology articles involved two important concerns. First, an effective procedure for searching among the articles in order to distinguish between nano-related articles and other articles had to be found. Second, an appropriate source of the journal papers which provides the most reliable and comprehensive data had to be determined. Each of the following sections describes the methods we used and the decisions we made in this respect. The final subsection then explains the extraction procedure.

4.1.1.1 Nanotechnology Keywords

Selecting nano-related articles in the databases of scientific papers is not a trivial issue. This is caused by the nature of nanotechnology, which is very multidisciplinary and thus it covers a wide range of nanotechnology disciplines, materials and systems. Moreover, at this point there is still no formal categorization in the databases of scientific articles, which would help us in finding the ones related to nanotechnology. Therefore, finding an effective and efficient search strategy which would include all the nanotechnology-related papers and exclude all the ones which are not has become a challenging issue. Most of the scholars use some sets of keywords in order to distinguish the nano-related articles from others, e.g. [2], [92], [93], [94], [95], [96], [97], [98], [99], and [100]. However, given the multi-disciplinary nature of the nanotechnology and its ill-defined boundaries it is very challenging to select the proper set of keywords. For example, some authors use a very simple strategy – they use “nano*” keywords to construct their database [101] [102]. In other words, they assume that an article is related to this discipline if and only if it has a keyword with nano- prefix in its title [102], and/or author-supplied keywords [101].

This method is however not very effective, because not only nano- prefix does not cover all nano-related keywords, but also it will find many completely unrelated articles. For example, some terms like “quantum wire” or “molecular sensors”, which do not start with “nano” prefix, are frequently used in nano-related articles, while some other keywords like “nanometer” are very general, and can results in nonrelated articles.

Many of the keyword strategies in the reviewed articles use the nano- prefix as a basic filter and improve it with Boolean search. In other words, they use Boolean algebra to exclude the general or unrelated terms that start with ‘nano’ [103], and/or subsequently include the related keywords without nano- prefix [2] [92] [93] [94]. Some authors, as for example in Porter et al. [95], performed a modular Boolean search and tested a substantial number of potential search terms to evaluate how specific their search is. Also, there are even more sophisticated studies that used various iterative techniques with relevance feedback to find patterns in a set of keywords [96] [104].

For the purpose of this research we studied thoroughly all the existing keyword search strategies in the literature and created our own combined collection of keywords based on seven different sources ([93], [95], [96], [97], [98], [99], [100]). The final set of keywords was consulted with nanotechnology experts, who omitted some redundant keywords and excluded irrelevant ones. The wildcard characters[11] were also omitted from the keywords and substituted by several keywords instead. (See Appendix I: List of Nanotechnology Keywords)

---

[11] A character that may be substituted for one (?) or any number of (*) possible characters called wildcard characters.

4.1.1.2 Source of Data

For collecting the information on nano-related articles, we first compared different digital libraries and online databases of scientific articles and abstracts in order to select the one(s) which would best suit our purpose. Some of the important factors which we needed to take into consideration in the selection of the data source are the diversity of fields, authors' affiliation and address information, numbers of articles we can retrieve, number of publishers which the database includes or the ability to search the full text.

As nanotechnology is an interdisciplinary field of research, the diversity of fields was a significant concern in our research. Hence, some databases like CABI[12] [105] or Pub Med[13] [14] [106] that focus on a specific range of disciplines are put aside.

Number of records, including journal articles, articles-in-press[15] and conference proceedings, is another factor playing an important role in the database selection. Although the size of the database is a significant factor, it is more important to see in what sense a database is large. For example, Web of Science[16] is a lager database than Scopus[17] in terms of articles published until 1996, but after 1996 Scopus offers a bigger number of articles added per year. And, in 2000 it gets slightly ahead of Web of Science. Therefore, we needed to take into consideration the period of time on which we want to focus. As nanotechnology is an emerging field of science, the focus of this research has to be on recent years.

In addition to digital libraries, some scientific search engines like Google Scholar[18], Microsoft Academic Research[19], and Scirus[20] search many scientific libraries and

---

[12] Mostly covers agriculture, animal sciences, health sciences, plant sciences, and natural sciences

[13] http://www.ncbi.nlm.nih.gov/pubmed , accessed May 7, 2012

[14] Mostly covers life sciences and biomedical topics

[15] Accepted articles that are available before its publication

[16] http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science , owned by Thomson Reuters, accessed May 8, 2012.

[17] http://www.scopus.com/home.url , officially named "SciVerse Scopus", owned by Elsevier, accessed May 8, 2012

[18] http://scholar.google.com , owned by Google, accessed May 7, 2012

[19] http://academic.research.microsoft.com , owned by Microsoft, accessed May 8, 2012

[20] http://www.scirus.com , owned by Elsevier, accessed May 8, 2012

databases, returning a wide range of results. However, these search engines have their own limitations. Google Scholar, for example, offers results of inconsistent accuracy and a confusing list with many dead links [107] [106]. Or, Microsoft Academic Search used to cover only computer science until June 2011[21]. According to Ford and O'Hara [108], Google Scholar provides the best coverage of these three databases.

Another feature that a scientific database or a search engine can provide is the accurate and comprehensive metadata. Metadata is the data about data. Number of citations, information about the authors and the system keywords added to authors' keywords are some examples of metadata. In this study, we need this type of information in order to build the co-authorship network. For example, we need to know an affiliation for each scholar in order to be able to distinguish the scientists from the US, Canada, and Quebec.

Table 1 shows a comparison of the most frequently used scientific databases. Among them we selected Scopus for the purpose of this study, because it covers a wide range of diverse fields of study and provides more metadata. The main disadvantage of Scopus was identified as the inability of performing the full text search in the database. Scopus allows only the keyword search within the article title, article abstract and the author name. Given the wide range of disciplines where nanotechnology found its applications coupled with different practices in terms of the abstract content in various journals and disciplines we realized that there is a need to perform a full text search in the database. The next section describes how this obstacle was overcome.

---

[21] http://academic.research.microsoft.com/About/Help.htm#6 , accessed May 8, 2012

Table 1 - Comparison of scientific databases and digital libraries

Source	# of Articles	Affiliation Information	Diversity of Fields	Full Text Search	Additional Information
Science Direct	9.5 M[22]	Provided	Wide	Allowed	-
Scopus	44.4 M[23]	Provided	Wide	Not Allowed	Citations (per year), Affiliation History, Additional Keywords Authors' information (interest, Affiliation History, etc.)
Web of Science	49.4 M[24]	Provided	Limited	Allowed[25]	Citations
Microsoft Academic Search	35.3 M[26]	Provided	Wide, but mostly Computer Science	Allowed	Citations (per year), Affiliation ID, Authors' information (interests, co-authors, etc.)
Scirus	440 M WebPages[27]	Provided	Wide	Allowed	-
Google Scholar	Theoretically all the web	Not Provided	Wide	Allowed	Citations

4.1.1.3 Data Gathering

In order to get benefits from the additional information that Scopus provides, and to still be able to search the full text of the articles, we develop a new data extraction methodology which involves a combined use of Google Scholar and Scopus. The main idea is to use the full text available search in Google Scholar with the help of software called "Publish or Perish[28]" and then to search these results in the Scopus database. In other words, we filtered the results of Google Scholar to get the metadata of the Scopus. This new idea has its own pros and cons described below:

Advantages of combining the Google Scholar and the Scopus:

---

[22] From http://www.info.sciverse.com/sciencedirect/about , accessed May 8, 2012

[23] As April 2011, http://www.info.sciverse.com/UserFiles/sciverse_scopus_content_coverage_0.pdf , accessed May 8, 2012

[24] As April 2011, http://wokinfo.com/realfacts/qualityandquantity/ , accessed May 8, 2012

[25] Depends on the Journal

[26] As September 2011, http://academic.research.microsoft.com/About/Help.htm#6 , accessed May 8, 2012

[27] The number of unique articles is unknown

[28] Harzing, A.W. (2007) Publish or Perish, available from http://www.harzing.com/pop.htm ,accessed May 8, 2012

1. Scopus assigns a unique ID to each author. This feature allowed us to avoid disambiguation of similar names. This would otherwise be a very problematic issue: as an example, we would need to find out whether John Newman from University of Toronto is the same John Newman from University of British Columbia, but he changed his affiliation because he moved. Or, is it a different person with the same name? By using Scopus we avoided extensive database cleaning and additional analysis in order to distinguish different people.

2. Moreover, in Scopus we can find all the historical affiliations of authors. Using this feature we recognize an author when exactly he/she changed his department or organization. It is important to know where an author was affiliated at each year when we are comparing the regions through the time.

3. Scopus provides indexed keywords – additional keywords that assigned to an article by Scopus – which let us to evaluate our searching queries and categorize the articles of nanotechnology. The keywords provided by the authors themselves are not always very consistent. This Scopus feature was thus very helpful for us.

4. The data in Scopus follows a similar structure that makes the process of data gathering easier and reduces the amount of irrelevant data.

5. Results in Google Scholar have inconsistent accuracy that can be reduced by filtering through Scopus.

Limitations of this method:

1. Google Scholar is a beta version and does not accept wildcards. This limitation forces us to provide a complete list of keywords without wildcards.

2. Google Scholar has a limit on number of queries to avoid spambots. Spambots are computer programs that spider the web automatically, usually for advertisement purposes. Consequently, the Publish or Perish program would be assumed as a

spambot and would be forced to stop working. Google normally provides an API[29] for scholars to avoid this limitation, but this is not the case for Google Scholar, because of it is only beta version. This limitation delayed in the process of data gathering.

3. Google Scholar and most of the search engines does not provide more than 1000 results per query, assuming that what the user wants should be in first 1000 results, and otherwise the search should be more specific. But, for the purpose of our research we cannot be more specific with our query because this would narrow down our results, and our objective is to get all the results, not a few specific ones. In order to solve this problem, we broke the result set into fractions of fewer than 1000. In order to do that, for each query, we needed to include or exclude a keyword so that we end up with two queries with lower number of results, which are without any interaction and which cover the original result set. In some cases up to 20 levels of fractions are applied to divide one set of thousands results to many sets of fewer than 1000 results.

Due to the above limitation the process of getting information from Publish or Perish cannot be automated, which required a difficult manual work.

In order to support this methodology I developed a program in Java that gets the results from Scopus by interpreting the Publish or Perish output. This program then searches the name of each article in the Scopus search engine and gets the page of the article and its authors. All the information available on these pages is parsed (read) by the program and stored in a MySQL database for further analysis. Figure 3 shows the entity-relationship diagram (ERD) of the articles database.

The process of data gathering was very long and we encountered many technical problems, for example access to Scopus was unavailable (Concordia University did not have a subscription to the database at that time), which forced us to use temporary

---

[29] Application Programming Interface

subscriptions or resources of other universities. Also, our program was banned by Scopus several times because of a high traffic which we created on their website, which forced us to slow down the extraction process and also divide the extraction process among different computers. Furthermore, the changes in the website of Scopus make problem for us, because, we need to update our extraction code to fit with their new website style.

Encountering all these limitations and problems, after more than three months, we extracted 748,251 nanotechnology articles, where for each of them we have its name, abstract, keywords, references, the information on the publication and the journal and the citation each article received each year, and for each of the co-authors we have their names, a complete history of their affiliations, numbers of articles, co-authors, references and citations per year. Not all the available information will be used in this thesis, but the created rich database has already served several other students in their own research. Moreover, the created extraction program will be further used for the extraction of information from the same or similar databases in future.

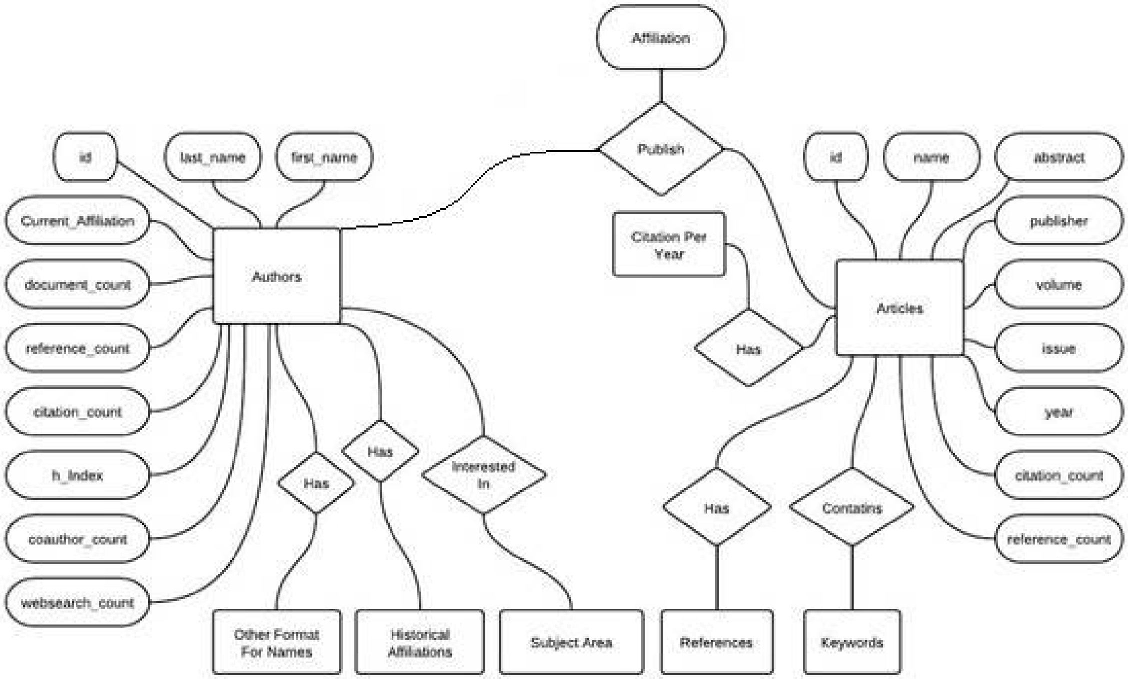


Figure 3 - Schematic entity-relationship diagram of articles database

4.1.2 Patents Database

To compare the Quebec, Canadian, and American nanotechnology collaboration networks, we need to know the geographical location of the residence for each inventor. United States Patents and Trademarks Office (USPTO) is the only database that provides this information, unlike the Canadian Intellectual Property Office (CIPO) database or the European Patent Office (EPO). In case of Canada and Quebec, using USPTO database instead of CIPO may cause a bias in the data, but considering the fact that Canadian inventors usually patent both in Canada and in the US, this bias is minimal [67]. Also, Canadian companies prefer to protect their intellectual property in the US, because of greater market opportunities and ease of accessibility [87].

In order to obtain the USPTO patent co-inventorship data, we use Nanobank database in this research. Nanobank is a public digital library consisting data on nanotechnology articles, patents and federal grants developed by Lynne G. Zucker and Michael R. Darby[30]. The patent database of Nanobank is based on the information from the USPTO and it includes 240,000 nanotechnology patents registered between 1976 and 2005.

According to Schiffauerova and Beaudry [87], this number is enormous compared to other similar databases. It was found that this database contains both nanotechnology relevant and “not so related” patents. Schiffauerova and Beaudry [16] used a cleaning strategy to obtain the largest possible extent of relevant data [87][31]. I used their database as the source of data for patents.

---

[30] Nanobank ©2007 by Lynne G. Zucker and Michael R. Darby, Los Angeles, CA: UCLA Center for International Science, Technology, and Cultural Policy and Nanobank. See the Nanobank database website: http://www.nanobank.org/.

[31] They used a strategy similar to what I did for articles in this research. The details about their strategy can be found in [87].

4.2 Data Manipulation and Cleaning

The patent database has been already cleaned by the authors - Schiffauerova and Beaudry -; however the article database needs to be cleaned. There are several sources of irrelevant or missing data. For example, some changes in the Scopus website, or some interruptions in the running of the program may cause the data to be lost or stored in inappropriate fields. Moreover, some irrelevant articles which are not related to nanotechnology may be collected by the Google Scholar. To solve these types of problems we performed the following cleaning processes:

- Empty field or fields with “strange” values (e.g. html tags), are identified and replaced by the correct values from the Scopus by another Java program created for this purpose.
- Non-English characters, mathematical notations, and some special characters that are stored with codes in html, were replaced by the original characters using MySQL queries.
- The complete text of each affiliation appears only in one field in Scopus. It was necessary to split the text into different fields, thereby allowing us to distinguish organization, department, city and country.
- Using the keywords that authors or Scopus assigned to the articles, potential irrelevant articles are identified and manually checked. Some of them were found irrelevant and deleted.

After the cleaning process of the database was complete, several pre-analyses were applied to the database to increase the performance of query executing. For example, for all fields of type character string, an index is created to increase the searching speed. Database indexes are data structures that increase the performance of information retrieval while also increasing the storage space.

4.3 Network Building and Analysis

4.3.1 Two-mode and One-mode Networks

In order to create the collaboration network of authors[32], the membership network is needed to be extracted from the database. The membership networks (also called affiliation networks[33], dual networks, or hyper networks) are networks with two types of entities: actors (authors) and events (articles). Because of these two types of entities, membership networks are also called **two-mode** networks [109]. Edges in these networks show the contribution of actor in events, i.e., an actor is connected to an event if he/she is the author of that article. Membership networks can be mapped on bipartite graphs in which we can divide the vertices into two disjoint independent sets. In membership networks vertices can be categorized into two groups (group of actors and group of events) in such a way that there are no edges connecting a pair of actors or a pair of events. In other words, every edge is joining an actor vertex to an event vertex [51]. In graph theory, graphs with this property called nondyadic [109].

A **one-mode** network is a network with one type of vertices that shows the connection between them using the edges. From each two-mode network, two one-mode networks can be extracted: network of actors and network of events. The network of articles[34] based on their mutual authors would not be very useful for the purpose of this research, and therefore we disregard this option. Instead, we are interested in the network of actors connected by the edges representing their mutual collaboration on an article (articles). In such a network, two vertices are linked to each other if and only if they have collaboration in at least one event (article co-authorship) [47].

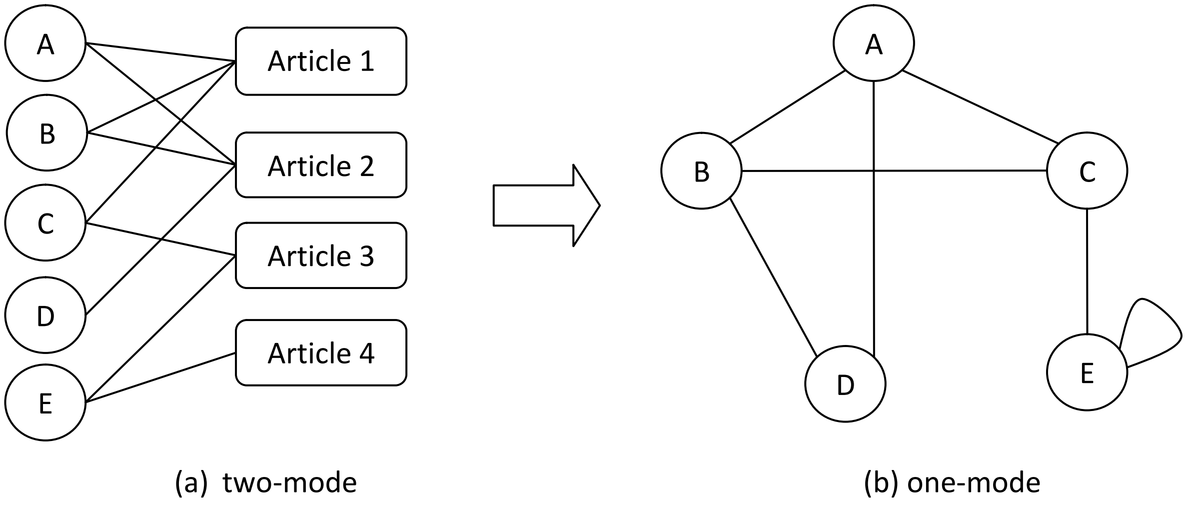
---

[32] In this section we explain the concepts and process using the terms “article” and “author”. The same scenario is valid for patents and inventors.

[33] The term “affiliation network” is the most common term for this type of networks. However, in this thesis, we used the same term for a subnetwork explained in 4.3.5. Therefore, we used the “membership network” to explain this concept.

[34] Hereafter, we use the term “network of article” for article-based collaboration network of scientists and “network of patents” for patent-based collaboration network of scientists for simplicity.

Figure 4 shows an example of the creation of a one-mode network, from a two-mode network. In this example, circles represent authors and rounded rectangles are articles. Here, we have four articles, in which Article 1 and Article 2 have each three contributors, Article 3 has two contributors and Article 4 has just one author. As explained above, in the two-mode network (Figure 4a), edges exist only between articles (events) and authors (actors), and there are no edges between authors or between articles. On the other hand, in one-mode network (Figure 4b) there is one type of vertices, here representing the authors, and edges show the collaboration of the authors. In this example, authors A, B, and C are connected because they are co-authors of Article 1, authors A, B, and D are connected because of they are co-authors of Article 2, authors C and E are connected because they are co-authors of Article 3 and, finally, author E has a loop, an edge with the same source and destination, because he/she is the only author of Article 4. In case of authors A and B, who both have two contributions, we can add multiple edges or we can assign weight to the edge, where the weight of the edge shows the number of collaborations between each two authors. This feature is omitted in the figure.



(a) two-mode

(b) one-mode

Figure 4 - Extracting one-mode network from two-mode network

4.3.2 Network Analysis Software

Building, analyzing and visualization of large complex networks cannot be performed in the absence of network analysis tools. There are many packages available for this purpose, both free and commercial. However, only few scholars compare them in terms of functionality, performance, support, and user friendliness. Huisman and van Duijn [110] studied 27 software packages, including 23 stand-alone programs and five utility toolkits from several aspects. They compare these tools in terms of data, functionality and support. For the data, they checked the type of data that software can handle (two-mode networks, large networks, etc.), the input format (adjacency matrix, edge/vertex input, etc.), and whether the software provides an option to indicate the missing values or not. From the functionality prospective they investigated the visualization and analysis that the package provides. And finally, in terms of support they mentioned if the software is free or commercial and whether a help and a manual exist for the software or not. Among these packages, they focus on three of the most general and well-known ones (UCINET, Pajek, NetMiner), and three with specific features (MultiNet, STRUCTURE, StOCNET) [110]. In our study, the most important feature is the ability of the package to deal with the large networks. According to Huisman and van Duijn [110], Pajek and MultiNet are the only ones which are capable to analyze such large networks as ours. On the other hand, some features like user-friendliness, which is an advantage of MultiNet, and NetMiner, is not an important factor for academic – non-commercial – applications [110]. We considered all the advantages and disadvantages of these packages, and selected Pajek for analysis and visualization of the collaboration networks in our research.

4.3.3 Time Events Networks

In order to study the network properties, two types of networks (based on two assumptions) can be created. First, the network of articles will be discussed, and then the similar argument will follow for the patents network.

In the first type of network, one can create a complete network including the articles from all the years. This method assumes that the collaboration between two co-authors in an article creates continuous connection between them. In other words, two individuals exchange information in that field long after their article is published. For example, it is presumed that two authors who published an article in nanotechnology in 2000 will still be in contact and still sharing their knowledge in 2012. This assumption thus does not distinguish whether a connection took place in 2000 or in 2012, which needs discussion. According to Newman [73], it is fairly probable that a pair of scientists who collaborate on a research leading to the publication of an article will create a social connection in terms of academia. However, he argued that this is not always true and it also depends on the field of study. For example in case of high-tech physics, which can overlap with nanotechnology in many cases, there are large collaborations between co-authors who have never even met each other [73], and thus probably do not remain in any contact after the publication of the article. Complete networks have obvious flaws, and thus we will work with them only marginally in this study.

In this study, we build a type of time event networks based on a second assumption, in which each collaboration pair considered to have an age. We assume that the knowledge exchange between two co-authors of one article does not persist beyond three years after the publication of the article. In other words, the age of an edge is three years and after that it is assumed that the authors are disconnected and no further knowledge is transmitted through them unless they publish an article again. Based on this assumption, we have created 13 sub-networks corresponding to three-year moving window. The first sub-network starts from 1994 to 1996, and the last one includes the articles between 2006 and 2008.

The same argument is valid for the patents database. According to Dahl and Pederson [111], who studied the informal contacts in industrial clusters, the relationships among co-inventors of a patent will continue even after they finish the project. They observed that patent co-inventors will be in social contacts, and share their knowledge.

For choosing the time interval which we will study, we consider several constraints. First of all, as nanotechnology is a relatively new emerging science, we omit the years before 1990s. Also, there are relatively few articles published and patents registered before 1990s. We clearly observe a significant difference in the number of articles before 1994 and after 1994. Another constraint is the Nanobank database, which is our source of data for patent information, which does not contain the patents after 2005. Talking all these into consideration, we decided to study the co-authorship network of articles from 1994 to 2008, and limit the co-inventorship network of patent into a shorter period from 1994 to 2004.

4.3.4 Regional Networks

As mentioned before, the aim of this study is to compare the innovation diffusion and knowledge transmission in Quebec, Canada, and the United States through the social network analysis tools. We will focus on two kinds of comparisons: First, we compare the network characteristic of Quebec with the ones in the rest of Canada (all provinces except Quebec) in order to make conclusions related to the differences and similarities of the innovation creation and the related policies in Quebec and in other provinces. Second, we will compare the Canadian nanotechnology collaboration network with the one of the United States in the context of the national innovation systems discussed previously. These two approaches will allow us to answer our research questions and validate the hypotheses.

Having in mind that the international network contains all the vertices we need to find a method how to create regional subnetworks in order to be able to analyze the network properties for each region separately. In the first approach, used for example by Schiffauerova and Beaudry [67] [87], a regional subnetwork involves all the articles[35] in which at least one of the co-authors resides or has an affiliation in that region, while all other co-authors for these articles are included. In this method, the indicators and network characteristics will be calculated just for that region ignoring the effect of other

---

[35] We explain for the articles and the same approach can be applied for the patents

authors in other regions. Studying a region network separately without considering external collaboration of the actors in that network can however bring significant bias to the results. For example, assume that authors A and B, both residing in Quebec, do not have any mutual collaboration (defined by an article co-authorship). However, there is edge between author A and another author C residing in Ontario, and similarly author B has a joint article with an author D in Ontario. Now, assume an author E, also residing in Ontario, has separate collaboration both with authors C and D. We also assume C and D are not connected. See Figure 5a for a better understanding of the situation. If we follow the first approach, when we are going to analyze Quebec regional network, A and B will not be connected to each other, while in the Canadian network, which will include all of the authors (A, B, C, D and E), there will be a path from A to B leading through their Ontario collaborators. In Canadian network it will thus be assumed that A and B share their knowledge, while in the Quebec network it will be assumed that there is no connection between them and no knowledge transmission can take place. See Figure 5b.

Using the second approach, we would first calculate all the network characteristics pertaining for all the vertices (authors) in the complete network which includes all the regions. Then we would identify the authors who reside in each region as the network members of that region. The advantage of this method is that we would consider the network effect of authors who reside outside the analyzed region. This is in fact a very important effect which should not be excluded, especially for the countries like Canada, which has a lot of collaboration edges with the United States. Excluding the effect of American collaborators would completely decompose the Canadian (or Quebec) subnetworks into separated components [112] and thus greatly influence the network characteristics. In Figure 5c, you can see that although authors C, D, and E are not considered as members of Quebec Network, their effect on knowledge transmission from author A to author B is considered. However, the network indicators calculated for each author will not be considered for authors C, D, and E, because they do not reside in Quebec.

Considering all the mentioned advantages of the second approach we have decided to follow the approach in this research.

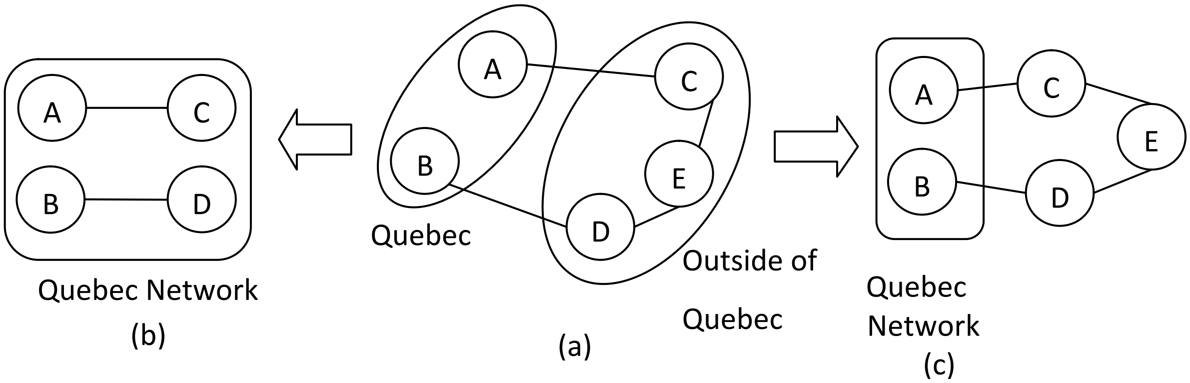


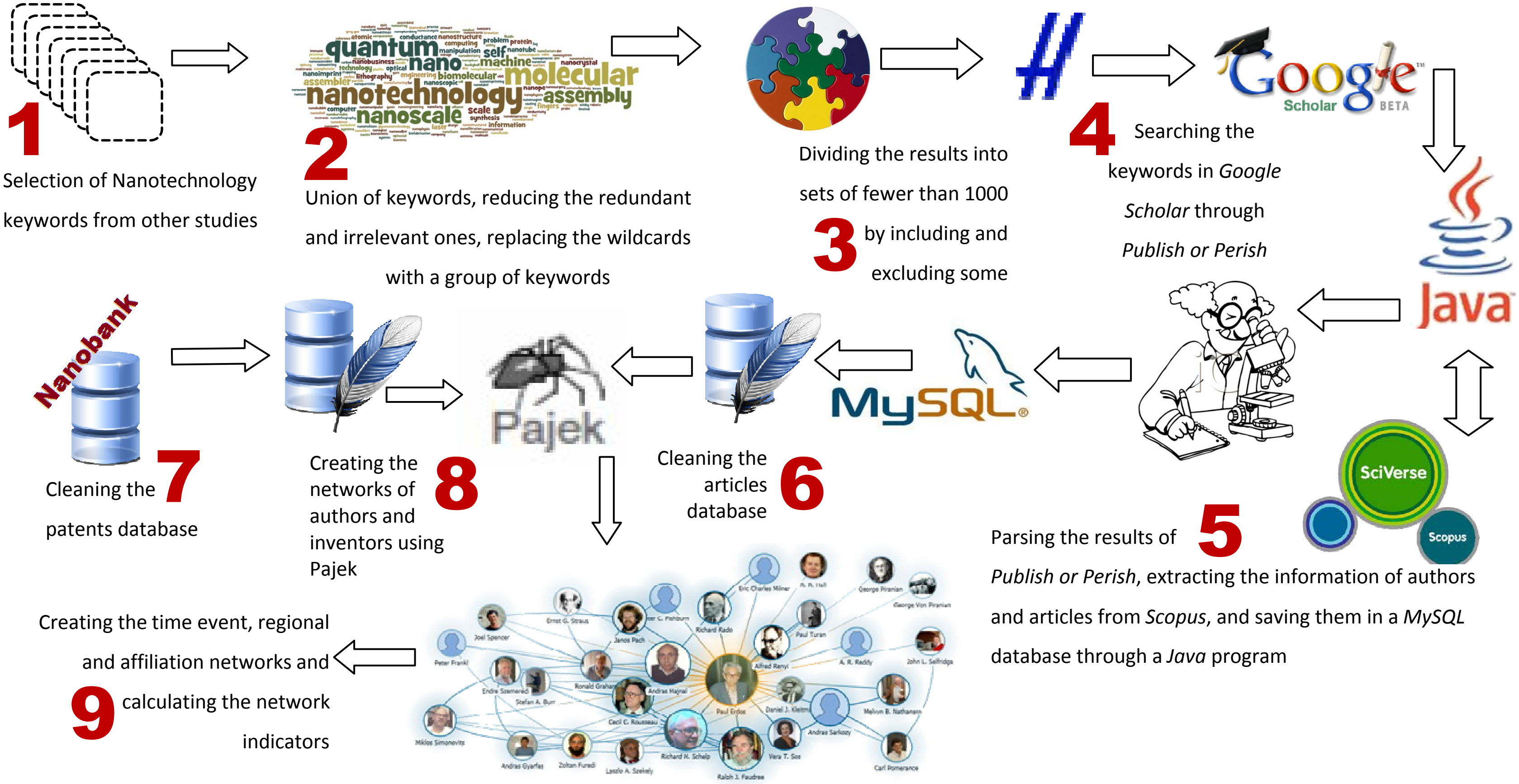
Figure 5 – Extracting the regional networks: (a) The geographical residence place of each actor. A and B reside in Quebec and the rest reside outside of Quebec. (b) Network of Quebec generated with the first method (ignoring the effect of actors in other regions). C and D are considered as actors of Quebec networks because they have collaboration with Quebec residents. However, there is no path between A and B. (c) Network of Quebec generated by the second method (considering the effect of actors in other regions). This time only A and B are assumed as a part of the network and C, D, and in not. However, there is a path between A and B which will affect the network properties.

4.3.5 Affiliation Networks

Another objective of this study is investigating the collaboration characteristics of scientists from academia and the ones from the industry. In order to distinguish between these two groups, we use the affiliation data that has been stored in the database. We assume that authors who affiliated to universities, schools, academies, colleges, and faculties as academic, and the rest of authors as non-academics. We are aware that this is a simplification, since some of the non-academic affiliations may be governmental research labs, non-for profit organizations, etc., but we found that the presence of such affiliations is quite minimal in our database. Therefore, hereafter the term non-academia and industry are interchangeable.

This categorization is not possible for the patents database, because this database includes the residence address of the patent assignees which does not show whether they are academics or non-academics. Therefore, our analysis on university-industry collaborations is limited to network of articles.

Figure 6 shows the summary of methodology steps from nanotechnology keywords to collaboration networks. In the following chapter, we will discuss network analysis results.



7 sets of keywords, reducing the
sets of fewer than 1000 (because
ngine; 5. Development of a Java
store them in relational MySQL
of patents to omit the not-nano-
ors and inventors deriving the 1-

I. The logos for Publish or Perish,
f the authors is the collaboration
n from http://findicons.com;

5. RESULTS

5.1 Fundamental Results on the Networks

In this study, two series of collaboration networks in field of nanotechnology have been constructed. The first group is based on the co-authorship of nano-related articles in 3-year intervals from 1994 to 2008, and the second one is based on the patent co-inventorship between 1994 and 2002 in 3-year time windows. The reasons for the selection of these periods and 3-year time frames are following: First, nanotechnology is still an emerging field, and the number of articles and patents in older data is much lower than in newer data. 1994 is chosen as the base year of the study, because there is a considerable difference in the number of researchers, articles and patents before and after this year. Second, the nanotechnology patent database used in this study is incomplete after 2002, which limits our time frame for this network. Third, because of the changes in the network in terms of entering and leaving of researchers and disappearing of their connections over time, we use a 3-year moving time window to examine the time evaluation of the networks.

In this section, we will present basic results related to the complete networks, while not taking into consideration the regional aspects. Also, we will discuss how social networks models can illustrate the characteristics of these two networks.

5.1.1 Characteristics of Article-based Networks

In Table 2, we find a list of basic characteristics of collaboration network of authors (created based on the co-authorship of scientific articles) for each period. The results in this table are categorized in the following 7 groups; (1) size of the network, (2) properties of collaboration pairs, (3) authors' collaboration pattern, (4) fragmentation measures, (5) structural cohesion, (6) distance indicators, and (7) clustering coefficient (cliquishness).

The table shows that the size of the collaboration network is growing over time in terms of number of authors (vertices of the graph), number of collaboration pairs (edges of the graph), and the total number of collaborations (sum of the weight of the edges in the graph).

According to these results, 13% to 18% of the collaborations occurred between authors who have already collaborated in that period. Also, the value of average number of collaboration per pair shows that on average, each pair of authors collaborates more than once. However, most of collaborations have not repeated again, as the median of collaboration per pair has value of 1. Surprisingly, there are some pairs of authors in each period of time that have more than 40 collaborations. This value in the moving windows of 2003-2005 and 2004-2006 reaches up to 81 repeated collaborations.

In addition to collaboration pairs, this table provides some indicators about each author in each period, such as average and maximum number of collaborators, and average and maximum number of collaborations. The average number of collaborator per author represents the degree of the vertex, which is also called the average degree centrality, while the average number of collaboration per author indicates the weighted degree of the vertex, which is obviously greater than or equal to average number of collaborators. Again, large values for maximum number of collaborators and collaborations show the existence of very productive authors in the network.

Moreover, the indicators that show the level of fragmentation, introduced in 2.2.1.2, are reported in this table. According to the results, the number of components is increasing as the size of the network grows. Also, the size and share of the largest component is growing which shows higher level of connectivity and lower level of fragmentation over the time. Other indicators like average component size and number (and percentage) of isolated authors are also reported for this network.

The other basic characteristic of the network is structural cohesion which refers to the degree to which vertices are connected to themselves. Density of the network is the most common indicator for structural cohesion. This indicator shows the proportion of edges over the maximum possible number of edges. For a network with n vertices, the maximum number of edges is $\frac{n\times(n-1)}{2}$. Density in complete graph, in which all vertices are connected to each other, is 1, and in a graph with no edges is 0. The results show that the density of the collaboration network is decreasing as the size of the network is increasing[36].

Furthermore, several important indicators reported in Table 2 are defined based on the geodesic distance of the authors. These indicators are calculated only in the largest component of the graph, because the distance is not defined in unconnected components. As defined in 2.2.1.2, the closeness centrality of a vertex shows how close that vertex is to other vertices. The closeness centralization of a network is defined as the variation of closeness centrality of vertices divided by variation in closeness centrality of a star network. Star network is a network in which one vertex is connected to all other vertices, and there are no edges between other vertices. Star networks have the maximum possible variation in closeness centrality [47]. The second indicator of this group is the average distance between reachable vertices, or degree of separation. This indicator shows that in the largest component of these huge networks, authors can be in fact reached only through few links. We will use this indicator in the section 5.1.3 to show how collaboration networks follow the network models explained in 2.2.1.1.

Finally, an indicator for cliquishness of the network is calculated for each period. Cliquishness in collaboration networks refers to the likelihood that two authors with a mutual collaborator also collaborate with each other. Network with high values of cliquishness tends toward dense local neighborhoods, in which individuals are better interconnected with one another [67]. The values of clustering coefficient, a common factor for cliquishness, are shown in the last line of this table. These values will also be used in 5.1.3 for network modeling.

---

[36] The values for density in the table are multiplied by 10^5.

5.1.2 Characteristics of Patent-based Networks

Same indicators as we calculated for the network of scientists are also calculated for the collaboration network of inventors, and the results are reported in Table 3. Comparing to the article-based network, the co-patenting network is smaller in terms of number of inventors (vertices), number of collaboration pairs (edges) and number of collaborations (sum of weights of edges). Also, this network is less central in terms of closeness centralization in all the common intervals from 1994-96 to 2000-2 (See Figure 7).

Similar to the collaboration network of scientific authors, the co-patenting network shows an increasing trend in terms of its size and the collaboration intensity, except the last period. The number of patent inventors, collaboration pairs, and total number of collaborations in fact decreased in the last period which happened because of the incompleteness of the database in our belief.

Table 2 - Basic characteristics of scientific authors' collaboration network (articles)

NETWORK OF ARTICLES	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
Number of Authors	89,626	100,089	109,654	121,564	128,668	141,551	158,573	178,890	200,595	231,315	269,683	309,570	355,767
Total Number of Collaborations	758,918	737,800	824,730	935,972	908,246	954,872	1,054,936	1,286,198	1,569,220	1,981,452	2,482,778	2,981,572	3,632,412
Number of Collaboration Pairs	303,893	288,640	318,581	360,004	346,856	370,205	411,656	498,552	598,367	738,473	919,726	1,112,209	1,382,925
Number of Repeated	40,049	43,586	49,158	56,378	56,517	57,085	62,110	77,239	98,473	129,927	164,718	195,845	227,811
% of Repeated Collaborations	13.18%	15.10%	15.43%	15.66%	16.29%	15.42%	15.09%	15.49%	16.46%	17.59%	17.91%	17.61%	16.47%
Avg. of Collaborations / Pair	1.2818	1.2674	1.2785	1.2996	1.2492	1.2090	1.1959	1.2685	1.3545	1.4612	1.5389	1.5803	1.6061
Median of Collaborations / Pair	1	1	1	1	1	1	1	1	1	1	1	1	1
Max of Repeated Collaborations	54	46	56	59	56	46	50	52	66	81	81	80	66
Avg. of Collaborators / Author	6.7814	5.7677	5.8107	5.9229	5.3915	5.2307	5.1920	5.5738	5.9659	6.3850	6.8207	7.1855	7.7743
Max of Collaborators / Author	373	241	216	229	231	189	186	213	252	245	244	247	315
Avg. Collaboration / Author	8.4676	7.3714	7.5212	7.6994	7.0588	6.7458	6.6527	7.1899	7.8228	8.5660	9.2063	9.6313	10.2101
Median of Collaborations of Authors	4	4	4	4	4	4	4	4	4	5	5	5	5
Max of Collaborations /Author	605	613	832	895	891	723	576	580	670	666	715	735	861
Number of Components	26,172	27,901	30,120	32,970	36,985	41,672	46,865	48,409	49,722	51,484	52,049	52,421	51,591
Size of Largest Component	41,662	52,052	60,908	68,929	69,747	75,000	82,534	97,702	117,143	143,673	178,031	214,152	254,283
Share of Largest Component	46.48%	52.01%	55.55%	56.70%	54.21%	52.98%	52.05%	54.62%	58.40%	62.11%	66.02%	69.18%	71.48%
Average Component Size	3.4245	3.5873	3.6406	3.6871	3.4789	3.3968	3.3836	3.6954	4.0343	4.4929	5.1813	5.9055	6.8959
Number of Isolated Authors	20,033	21,736	24,244	26,699	30,200	34,019	38,309	38,933	39,974	41,041	40,475	39,562	3,670
Percentage of Isolated Authors	22.35%	21.72%	22.11%	21.96%	23.47%	24.03%	24.16%	21.76%	19.93%	17.74%	15.01%	12.78%	10.31%
Network Density ($\times 10^5$)	7.57	5.76	5.30	4.87	4.19	3.70	3.27	3.12	2.97	2.76	2.52	2.32	2.18
Closeness Centralization of Largest Component	0.1467	0.1382	0.1353	0.1348	0.1304	0.1262	0.1216	0.1270	0.1307	0.1267	0.1280	0.1327	0.1323
Avg. Distance of Reachable Vertices (Degree of Separation)	5.4023	5.8204	5.9246	5.8549	5.7769	5.8669	5.8813	5.7991	5.8262	5.7788	5.7741	5.8177	5.7824
Clustering Coefficient	0.8611	0.8549	0.8474	0.8453	0.8502	0.8514	0.8580	0.8535	0.8461	0.8384	0.8335	0.8301	0.8294

Table 3 -Basic characteristics of scientific inventors' collaboration network (patents)

NETWORK OF PATENTS	94-96	95-97	96-98	97-99	98-00	99-01	00-02
Number of Inventors	81,174	89,235	93,230	98,999	101,946	103,491	95,925
Total Number of Collaborations	445,168	506,374	509,692	565,960	618,330	664,486	628,068
Number of Collaboration Pairs	150,383	171,370	179,901	200,184	214,351	228,197	216,735
Number of Repeated Collaborations	38,124	43,003	40,123	43,380	48,420	52,032	48,635
Percentage of Repeated Collaborations	25.35%	25.09%	22.30%	21.67%	22.59%	22.80%	22.44%
Average Number of Collaborations per Pair	1.334	1.3739	1.3276	1.3704	1.4193	1.4756	1.4972
Median of Collaborations per Pair	1	1	1	1	1	1	1
Maximum Number of Repeated Collaborations	55	75	102	130	83	81	113
Average Number of Collaborators per Inventor	3.7052	3.8409	3.8593	4.0442	4.2052	4.4100	4.5188
Maximum Number of Collaborators per Inventor	61	65	69	78	82	97	104
Average Collaboration per Inventor	5.4841	5.6746	5.467	5.7168	6.0653	6.4207	6.5475
Median of Collaborations of Inventors	3	3	3	3	3	4	4
Maximum Number of Collaborations per Inventor	294	440	583	732	577	644	667
Number of Components	25,496	26,484	27,608	27,075	25,946	23,389	19,842
Size of Largest Component	16,824	20,876	21,905	26,280	30,579	33,579	31,648
Share of Largest Component	20.73%	23.39%	23.50%	26.55%	30.00%	32.45%	32.99%
Average Component Size	3.1838	3.3694	3.3769	3.6565	3.9292	4.4248	4.8344
Number of Isolated Inventors	14,687	14,844	15,401	14,393	12,957	10,114	7,082
Percentage of Isolated Inventors	18.09%	16.63%	16.52%	14.54%	12.71%	9.77%	7.38%
Network Density ($\times 10^5$)	4.56	4.3	4.14	4.09	4.12	4.26	4.71
Closeness Centralization of Largest Component	0.0654	0.0684	0.0797	0.0767	0.0706	0.0707	0.0637
Average Distance of Reachable Vertices (Degree of Separation)	5.0776	5.7005	5.2125	5.7024	6.3079	6.2121	6.1768
Clustering Coefficient	0.88417	0.88062	0.88292	0.88083	0.87925	0.8773	0.87854

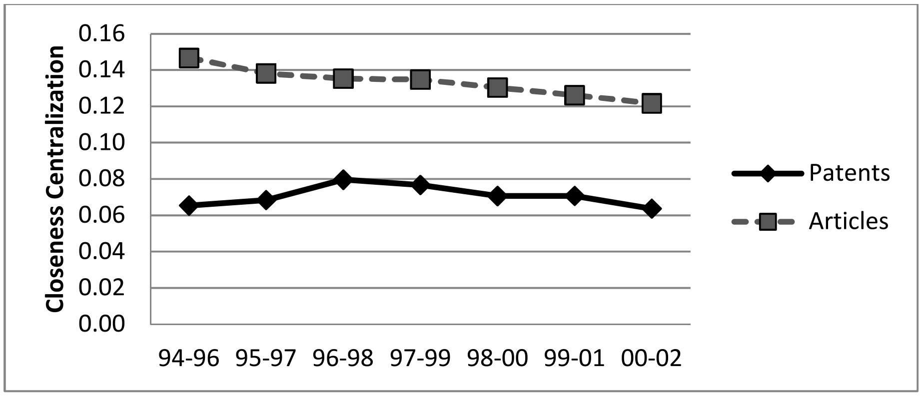


Figure 7 - Comparision of closeness centralization in collaboration network of authors (articles) and inventors (patents)

5.1.3 Network Modeling

As explained in 2.2.1.1, there are three classic models for complex networks; the **random networks model** developed by Erdős and Rényi [113], the **small-world model** introduced by Watts and Strogatz [59], and the **scale-free model** presented by Barabási and Albert [46]. In order to identify if these models fit our networks, we need to examine whether the properties of our networks follow the model properties or not.

The Erdős and Rényi random network model is distinguished by degree distribution. According to their model, if edges in a network are assigned randomly, most vertices have roughly the same number of edges, and the degree distribution follows the Poisson distribution for large number of vertices. In other words, a network is called random if it is exponentially rare to find an author who deviates from the average by having considerably more or fewer collaborations than the average author [46].

According to our results, the degree distribution in both networks in all periods does not follow the Poisson distribution, which implies that the linkages in the collaboration networks are not created randomly. Figure 8 and Figure 9 shows two sample degree distributions for each of the networks. As we can see, the degree distribution does not follow the Poisson distribution.

According to Watts and Strogatz model, complex networks with small-world effect should be clustered and have small degree of separation. Clustering coefficient and average distance of vertices are the indicators that illustrate these properties. According to the Table 2 and Table 3, the clustering coefficients in all studied networks are relatively high (in the range of 0.82 to 0.88), comparing to random networks which have values in the range of 10^{-5} for clustering coefficient [46]. Also, the average degree of separation is very close to other networks with small-world effect[37]. In conclusion, our collaboration networks show the properties of the Watts and Strogatz small-world

---

[37] Examples of average degree of separation are some other studied complex networks: species in food webs: 2; molecules in the cell: 3; the Web holds the absolute highest record of 20 to 22; the Internet (a network of hundreds of thousands of routers): 10 to 12 [46].

model. This result confirms our Hypothesis 1a that the collaboration networks follow the small-world phenomena.

Finally, the scale-free model of Barabási and Albert cannot perfectly simulate the collaboration networks. According to these models, degree distribution in scale-free networks follows the Power Law distribution [56]. However, in our networks, the number of authors with one collaborator (degree of 1), is fewer than authors with 2, 3, or even 4 collaborators. Therefore, we cannot claim that collaboration networks of scientists and inventors follow the degree distribution of scale-free network. This difference does not confirm or reject our Hypothesis 1b that collaboration networks are scale-free. The difference between our results and results of Barabási et al. [56] can be explained by the size of time windows.

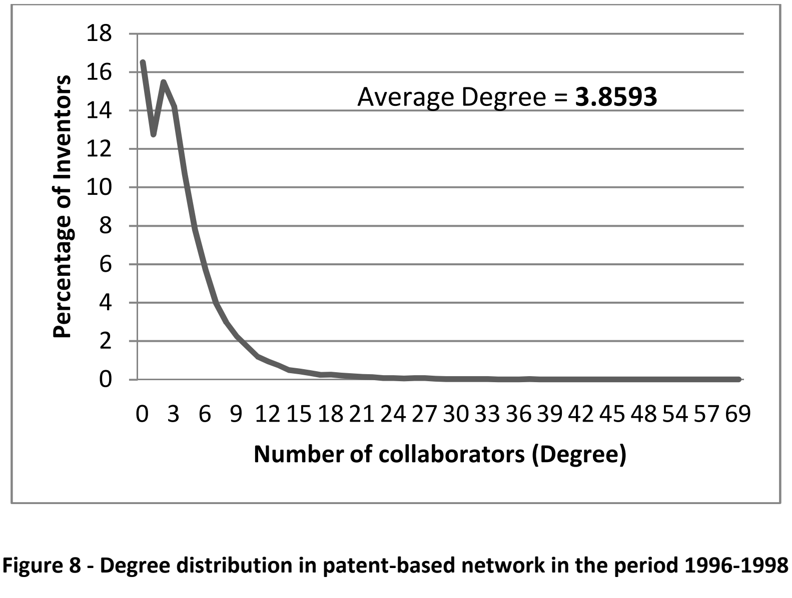


Figure 8 - Degree distribution in patent-based network in the period 1996-1998

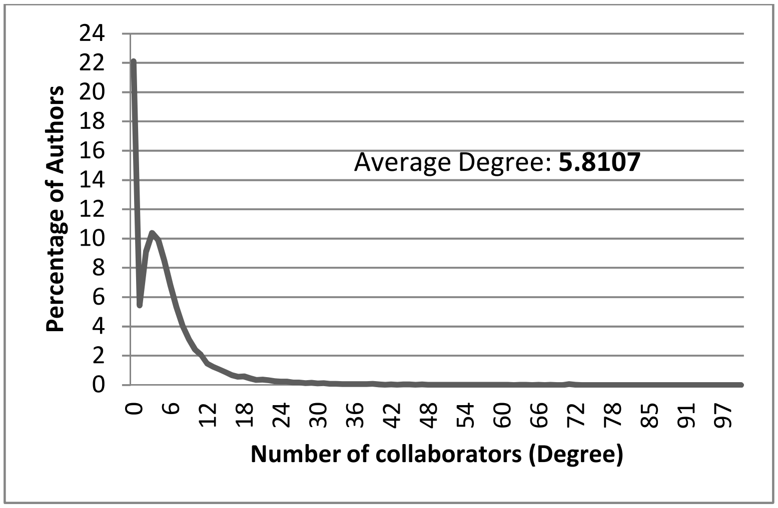


Figure 9 - Degree distribution in article-based network in the period 1996-1998.

5.2 Regional Characteristics of the Network Structure

After measuring the basic characteristics of the collaboration networks, we focus on the comparison of the network properties for the authors and inventors in Quebec, Canada, and the United States. In the co-authorship network of articles, authors of each region are distinguished based on their affiliation. For authors with multiple affiliations, we consider their first affiliation in the article as the main affiliation. On the other hand, the inventors in co-inventorship network of patents are allocated to the regions (i.e. Quebec, Canada or the US) based on the place of the inventor's residence. This is a consequence of the data available in the USPTO database, where only the residence address is provided, but the affiliation of each inventor is omitted. In this section, we perform two comparisons based on the network characteristics. First, the Canadian networks will be compared to Americans'. Second, the network of Quebec will be compared with the network of the rest of Canada.

5.2.1 Nanotechnology Collaboration Networks in Canada and in the United States

Comparison of Canada and the United States in terms of size is not very meaningful, because, the United States has expectedly a much bigger share in both patents and article networks. However, in order to compare the productivity of the authors or inventors, we can measure the average number of deliverables (article or patent) per individual. Figure 10 shows the number of patents per inventor. According to our results, the Canadian inventors are more productive than American ones in terms of patent registrations. In fact, an average Canadian inventor produced almost twice more patents than an average American one in the last period measured (2000-2002). This shows that there is really a huge research potential concentrated among a relatively small number of Canadian nanotechnology inventors. As the previous results have shown [16], these highly patenting Canadian inventors usually work for the American-owned companies. The fruits of their research efforts are thus collected by the United States. Regarding the co-authorship network, American authors produce, on average, more publications than Canadians (see Figure 11), but these results here are very close to each other, so the difference between them is very small.

It can be observed that there is considerable difference between values of average number of article per author and average number of patent per inventor. This is probably caused by a relatively small size of the patenting community, coming mainly from companies whose R&D researchers are focused on specific nanotechnology research, whereas the article-based network involves an enormous pool of academicians belonging to diverse fields of study and publishing in various other research domains as well.

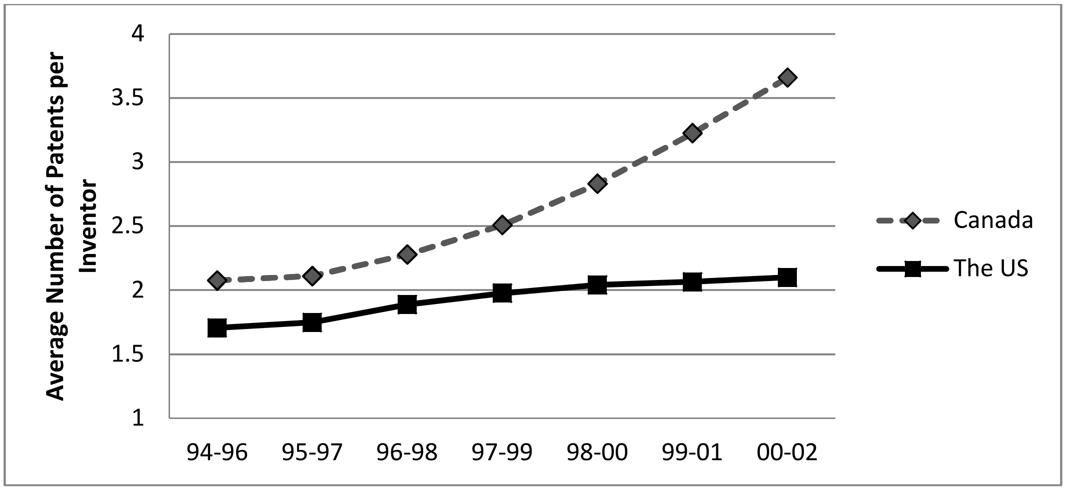


Figure 10 - Average number of patents per inventor in Canada and the US

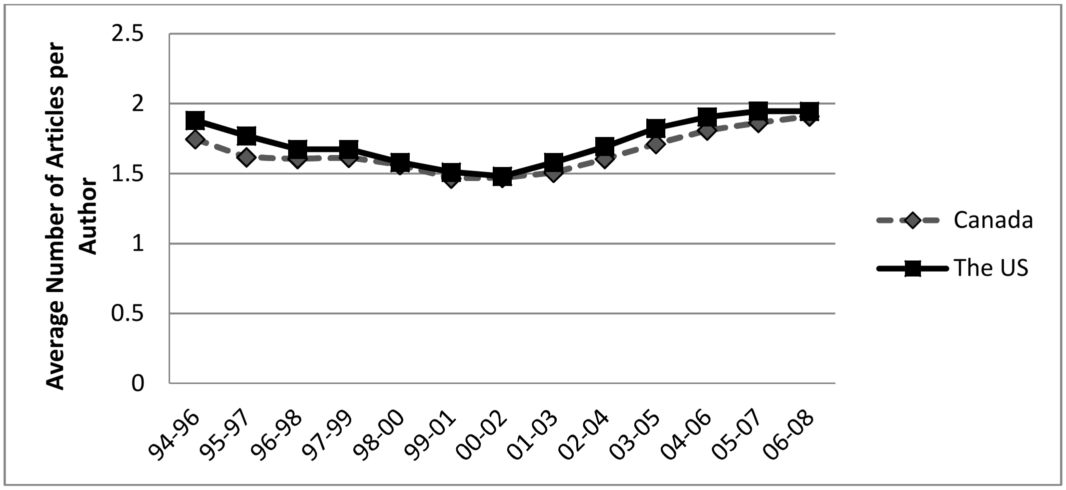


Figure 11 - Average number of articles per author in Canada and the US

In order to compare the role of American and Canadian scientists and inventors in each network, several indicators are calculated, such as average betweenness centrality, average degree centrality, average clustering coefficient, and average number of collaborators per article/patent.

In the co-authorship network of articles, our results show higher centrality of the American authors in terms of degree centrality and betweenness centrality almost in every period[38]. The higher degree centrality indicates that on average each American scientist has more collaborators than a Canadian one, while betweenness centrality shows that it is more probable that knowledge transfers through the American scientists to the rest of the network, comparing to Canadians'. These results indicate that American scientists have more important role in knowledge diffusion than Canadians' in the field of nanotechnology, and they have more central positions in the whole network. Figure 12 shows the comparison of the average betweenness centrality in the US and Canada, and Figure 13 compare the average degree centrality in these two countries.

---

[38] The only exception is the betweenness centrality in the first period.

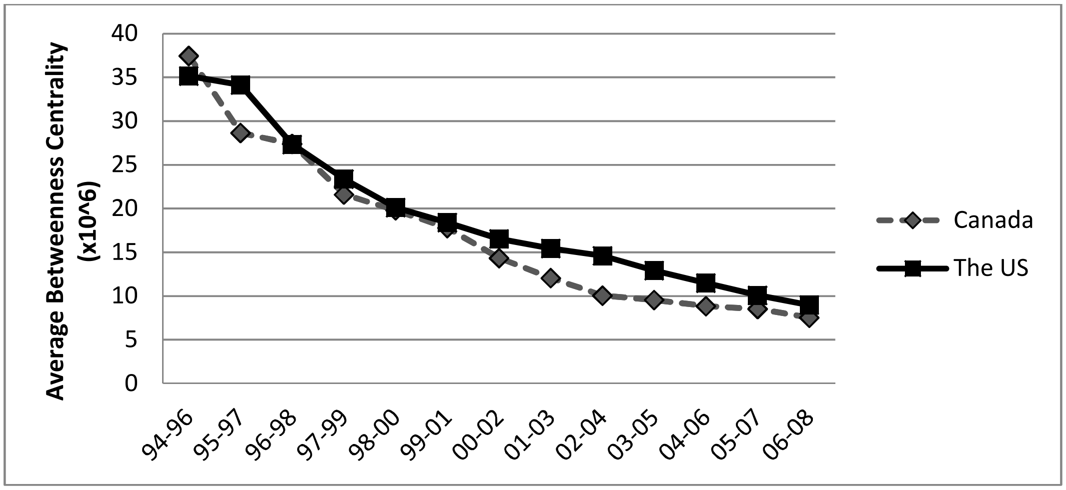


Figure 12 - Average betweenness centrality in the US and Canada in the network of articles (x10^6)

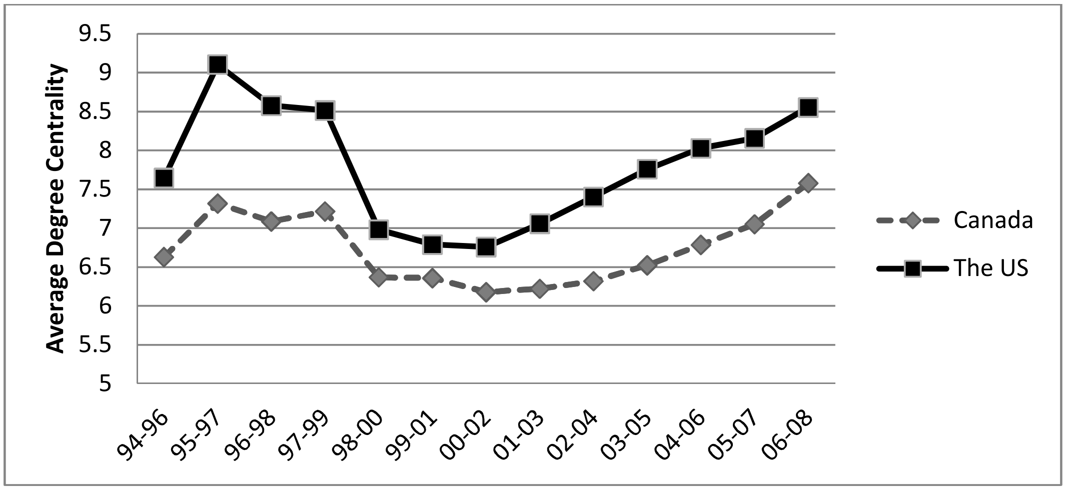


Figure 13 - Average degree centrality in the US and Canada in the network of articles

The results for the patent network are not exactly the same. Although, the results for the average degree centrality support the results from the other network, the results for the average betweenness centrality are not the same. This means, the American inventors have more collaborators in the patent network, however, their role in knowledge transmission is not more significant than Canadians' in all the periods. However, the differences in average betweenness centrality for Canadian and American inventors are not significant and the values are often very close to each other (see Figure 14 and Figure 15.)

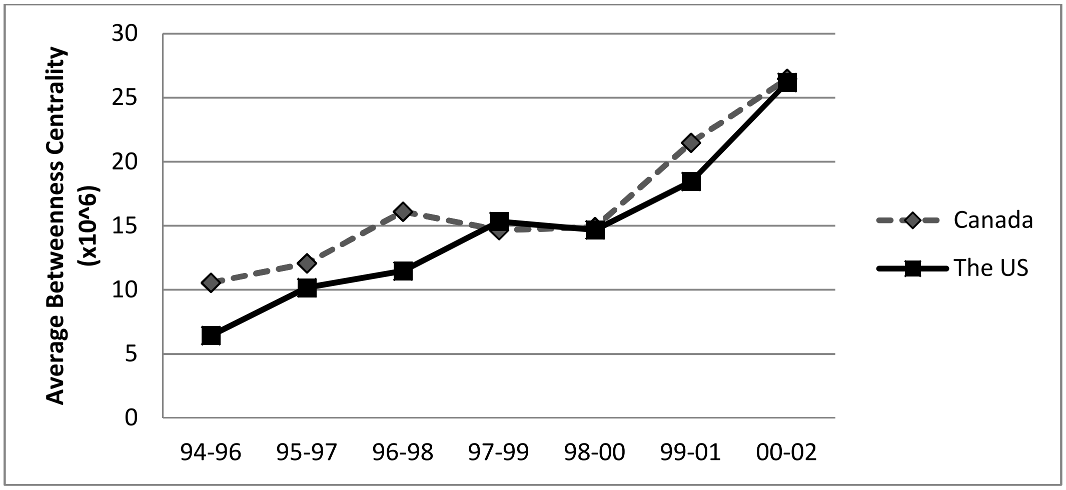


Figure 14 - Average betweenness centrality in the US and Canada in the network of patents (x10^6)

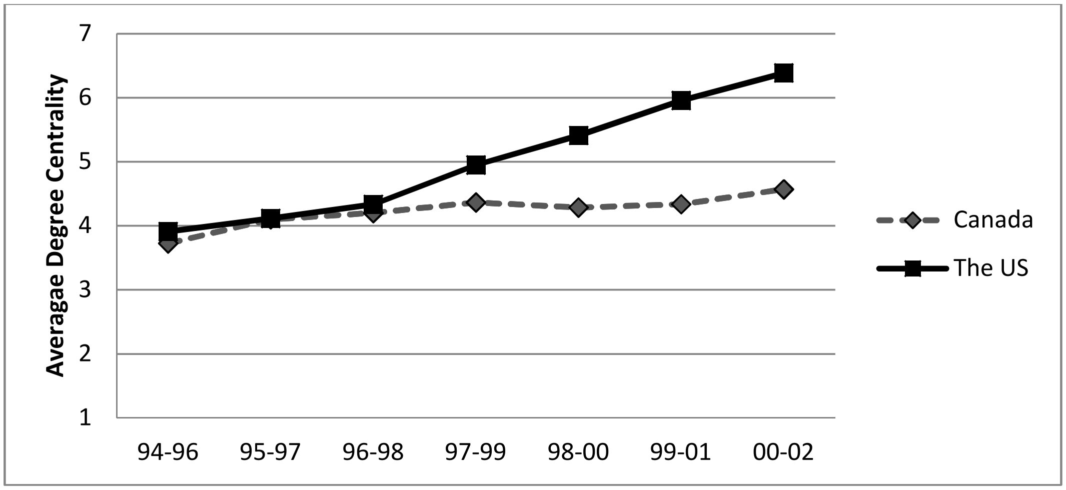


Figure 15 - Average degree centrality in the US and Canada in the network of patents

In order to compare the cliquishness of the American and Canadian networks, we calculate the average clustering coefficient of authors and inventors for each region and period. The results show that the Canadian inventors are more clustered compared to American ones (Figure 16 and Figure 17.) This means that Canadian inventors are more likely to be found working in highly clustered local neighborhoods.

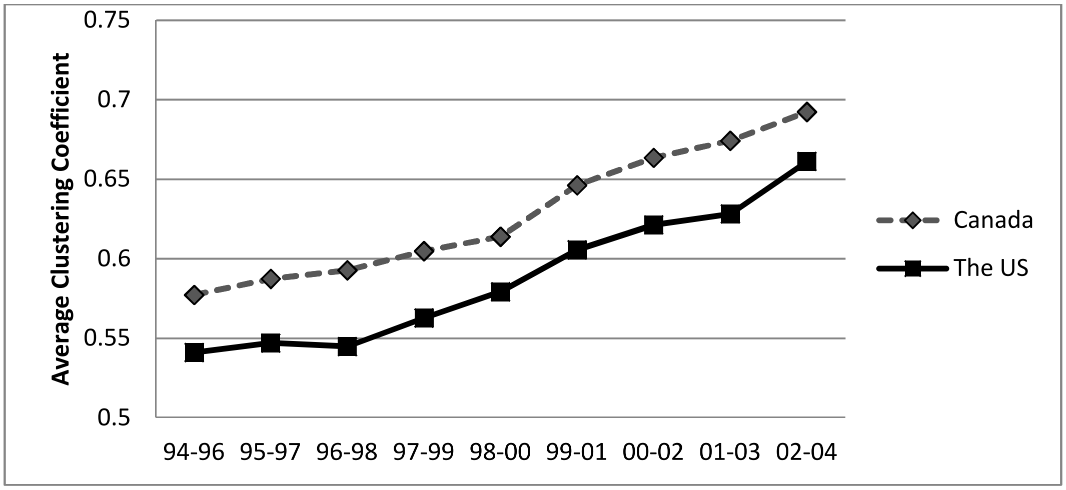


Figure 16 - Average clustering coefficient in the US and Canada in the network of patents

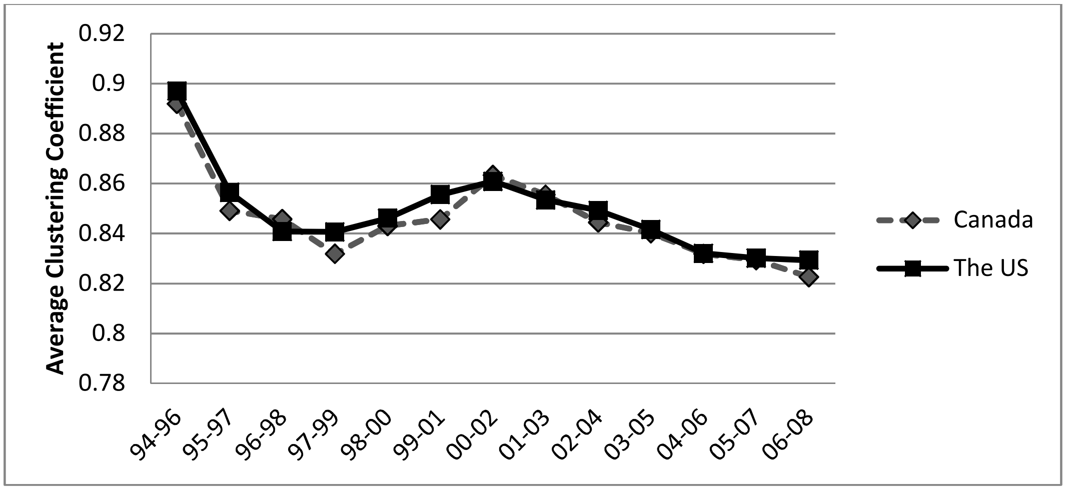


Figure 17 - Average clustering coefficient in the US and Canada in the network of articles

The co-authorship network shows different results. Cliquishness here is fairly comparable for both Americans and Canadians. The interesting fact in this network is the trend of clustering coefficient in these two regions. In late 1990s, both American and Canadian scientific authors are less clustered, and in the early years of 21st century, a more clustered network is observed. Again, after three periods, a decreasing trend is happened in both networks. On the other hand, quite opposite trend is observable in the co-patenting network in most periods.

The values for the clustering coefficient in the patent network are much less than the values for the article network. These results indicate that the scientists who collaborate in the article publication created more dense clusters than the inventors who collaborate in patent registration. There are two possible explanations for this finding: First, the scientific authors collaborate more with their friends of friends, and create triangles (three authors who collaborate with each other) in the network. The second possible explanation is that the average number of co-authors who collaborate in one article is higher than the average number of co-inventors who collaborate in one patent. If the researchers work in larger teams it is expected that their network would show higher cliquishness. We compared the patents and articles in terms of average number of collaborators per patent and article to find out who works in larger teams. Figure 18 and Figure 19 show these values for both the United States and Canada. As we can see, on average there are more collaborators in each article or patent in the United States compared to Canada. However, the average number of collaborators does not differ significantly in patents and articles. These results prove that the higher cliquishness in the article network is not caused by the existence of larger teams of article co-authors, but it is more likely due to the structure of the acquaintances in the academic world, where researchers working on very similar topics know one another very well. The collaborating “clusters” are thus formed regardless their various university affiliations. The acquaintances of researchers in industry are, on the other hand, more constraint by the companies’ boundaries, which are more rigid due to the obvious proprietary considerations and market rivalry. The results presented in this section accept all assumptions mentioned in Hypothesis 2, except the higher cliquishness for American scientists.

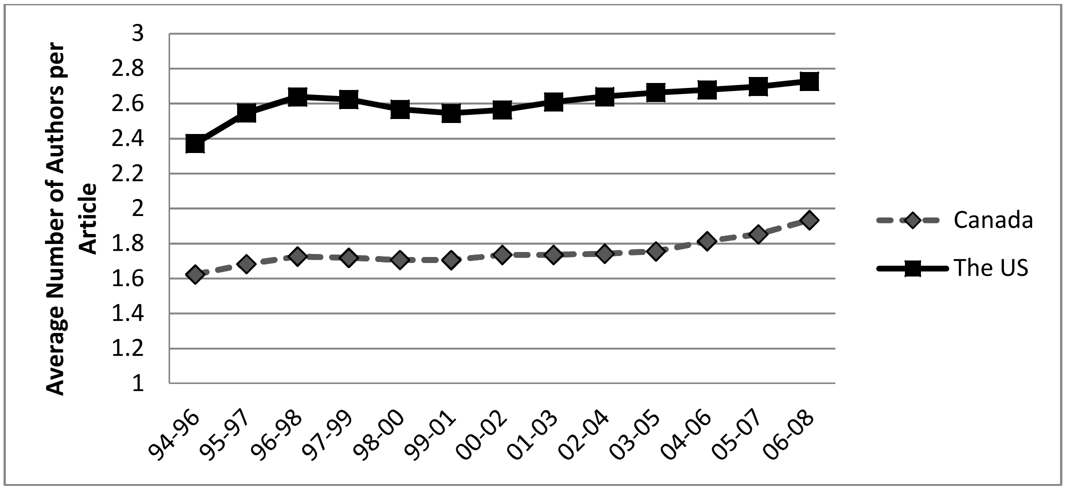


Figure 18 - Average number of authors in each article in the US and Canada

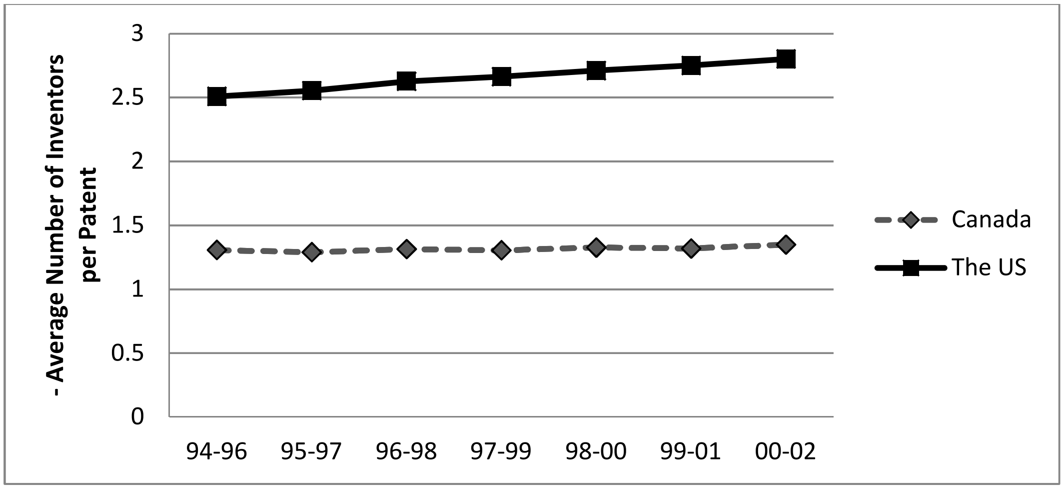


Figure 19 - Average number of inventors in each patent in the US and Canada

5.2.2 Nanotechnology Collaboration Networks in Quebec and the Rest of Canada

The same calculations are performed to compare the network of Quebec with the rest of Canada. In all the studied periods, the authors in Quebec are less central than the rest of Canada in terms of betweenness centrality (Figure 20 and Figure 22). However, their centralities have been slowly getting closer to the Canadian values, and we can say that the results in the last years for which we have the data are quite comparable with the rest of Canada, for both article-based and patent-based networks. Regarding the degree centrality (Figure 21 and Figure 23), even though the results shows that Quebec researchers always have a slightly less significant role in knowledge transfer in the network, their values are very close to the Canadian ones. On the other hand, Quebec inventors are more active in terms of the number of collaborators than the inventors in the rest of Canada. Their average number of collaborators is higher than the number of collaborating partners in other provinces.

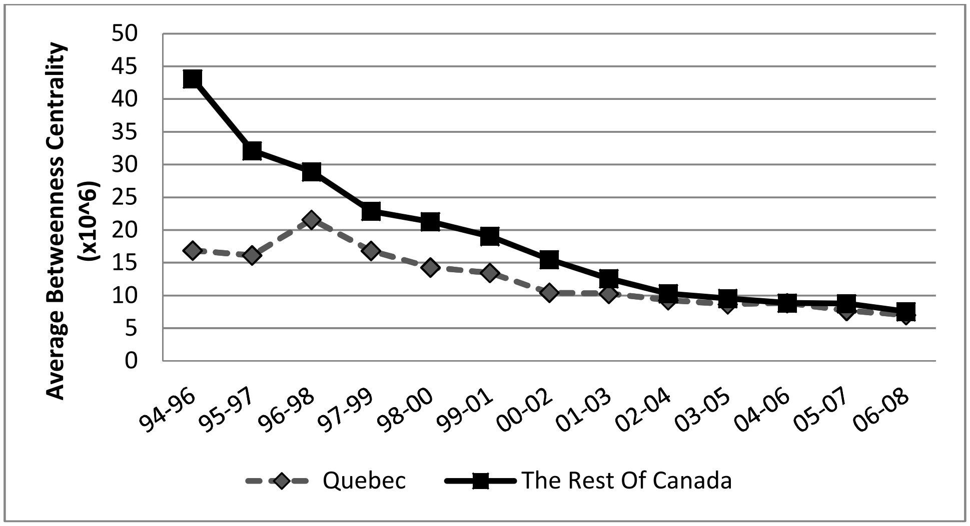


Figure 20 - Average betweenness centrality in Quebec and the rest of Canada in the network of articles ($x10^6$)

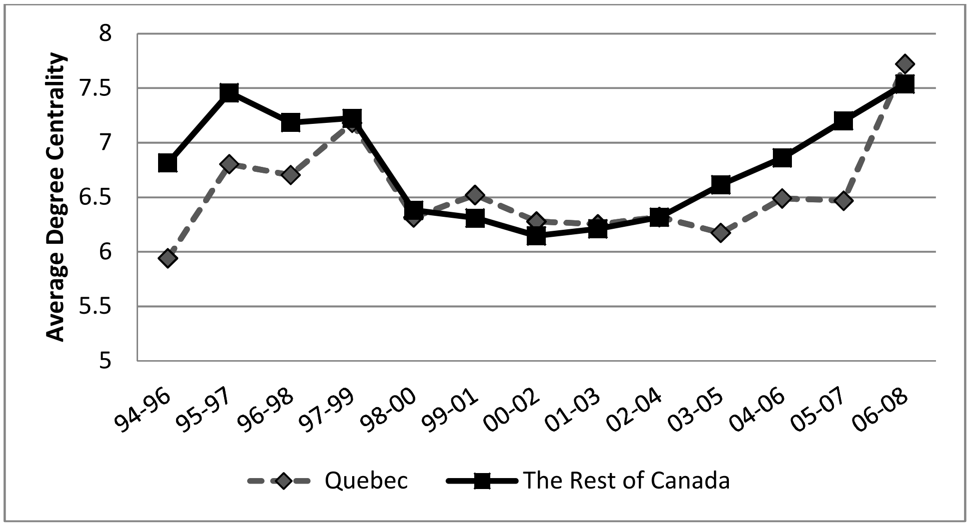


Figure 21 - Average degree centrality in Quebec and the rest of Canada in the network of articles

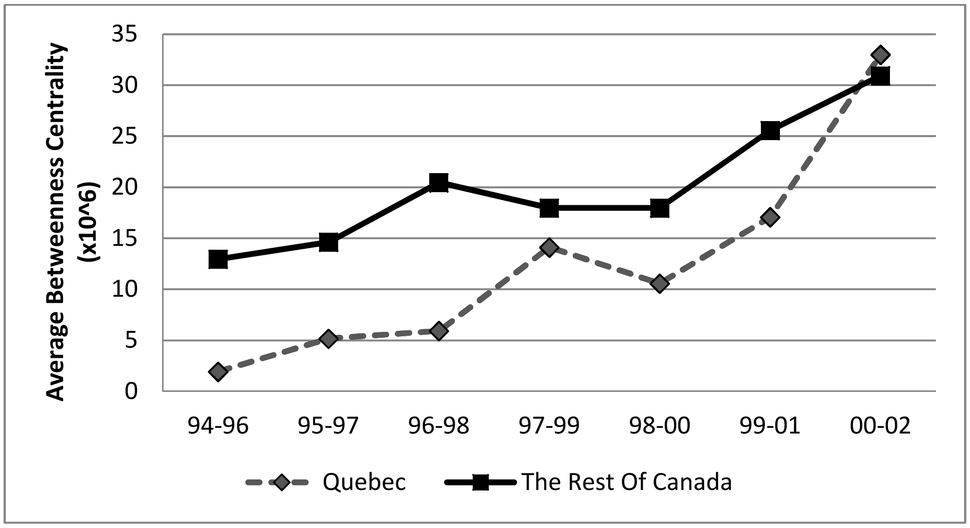


Figure 22 - Average betweenness centrality in Quebec and the rest of Canada in the network of patents

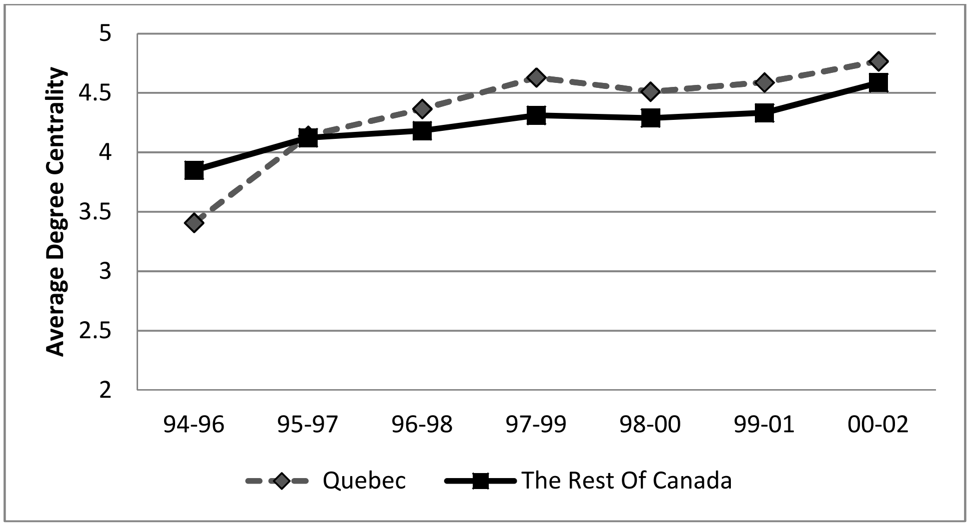


Figure 23 - Average degree centrality in Quebec and the rest of Canada in the network of patents

The clustering average coefficient is also calculated for the both networks, in each region and period (Figure 24 and Figure 25). According to the results of networks of articles, not significant difference between Quebec and the rest of Canada is observed. However, Quebec sub-network is more clustered than the rest of Canada in the patent network. Also, our results show an increasing trend in cliquishness of all Canada. Again, comparing the patent and article networks, we can see more cliquishness in article network and an opposite trend, i.e. cliquishness slightly decreasing with time in the network of article co-authors.

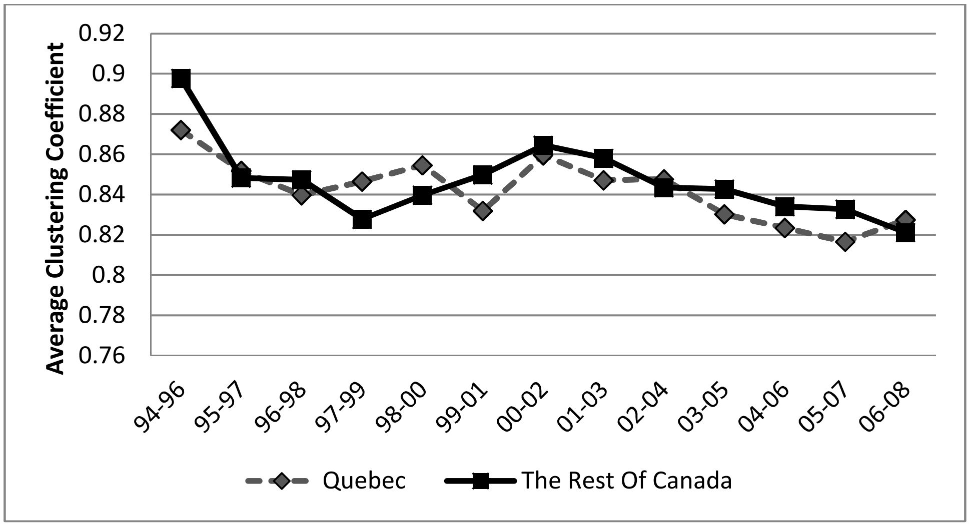


Figure 24 - Average clustering coefficient in the Quebec and the rest of Canada in the network of articles

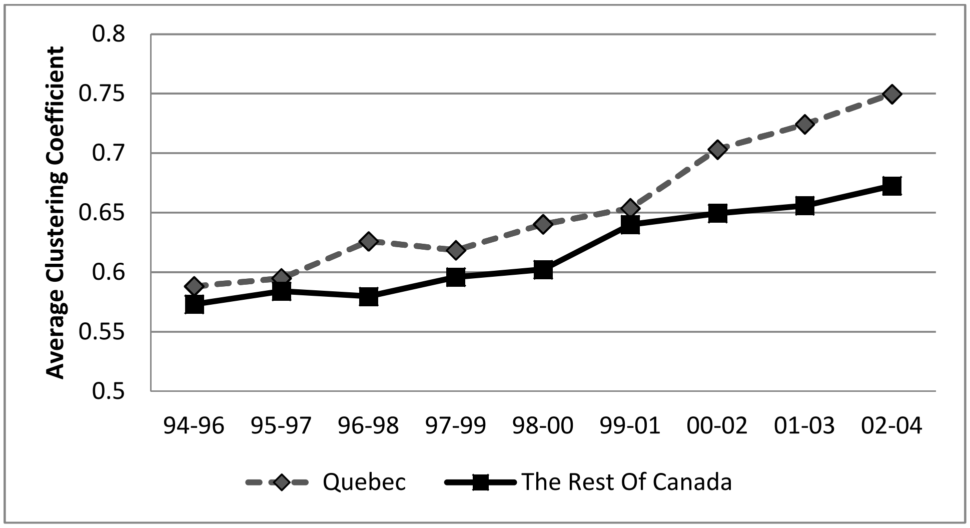


Figure 25 - Average clustering coefficient in the Quebec and the rest of Canada in the network of patents

5.2.3 Collaboration Patterns in the Articles-based Networks

From hereafter, our focus will be on the article-based collaboration networks. There are two reasons for this focus. First, the patents data is limited and we cannot examine the trends on few periods. Also, for creating the affiliation subnetworks (academia and industry), we need to know the affiliation of scientists, which is not provided in our patents database. In other words, based on the residence place of a patent assignee, we could not categorize them into academics and non-academics.

In order to compare the collaboration patterns in each region, the vertices of the network were categorized into four groups; Quebec, rest of Canada, the United States, and the rest of the world. Using this partitioning, the collaboration patterns inside each region and between every pair of regions are found. The following indicators are calculated for each region:

1. Sub-network size: share of each region in the whole network in terms of number of authors
2. Collaborations: number of collaboration pairs (edges), total number of collaborations (sum of the weights of the edges), average number of collaborations per author (average weighted degree), and maximum number of collaborations for an author (the maximum weighted degree)
3. Collaborators: average and maximum of collaborators per author (average and maximum degree)
4. Repeated collaborations: count of repeated collaborations, average and maximum of repeated collaborations per author
5. Isolated authors: count and percentage of authors who does not collaborate with other authors in that region
6. Connected component: size of the two first largest components and their share in the sub-network, average component size and diameter in the largest component

Similar indicators also calculated for each pair of regions. The results are provided in detail in Appendix II: Internal Collaboration Patterns of Regions in the Network of Articles and Appendix III: Collaboration Patterns between Regions.

According to our results, international (mainly American) collaborations form a significant part of overall collaboration pattern in Canada. On the other hand, American researchers tend to have more national collaborations. Figure 26 illustrates the summary of average collaboration of Canada and the United States in all the periods. Our results show only 30% of Canadian collaborations are carried out at the national level, while there are almost as much collaboration between Canada and the US (27%). Out of all Canadian international collaborations, 38% are directed towards collaborative partners in the United States. American researchers, on the other hand, tend to collaborate more within their own country, as 54% of American collaborations are held inside the United States. This was expected, since a great amount of research in various domains is performed within the US borders, and it is thus much easier for an American

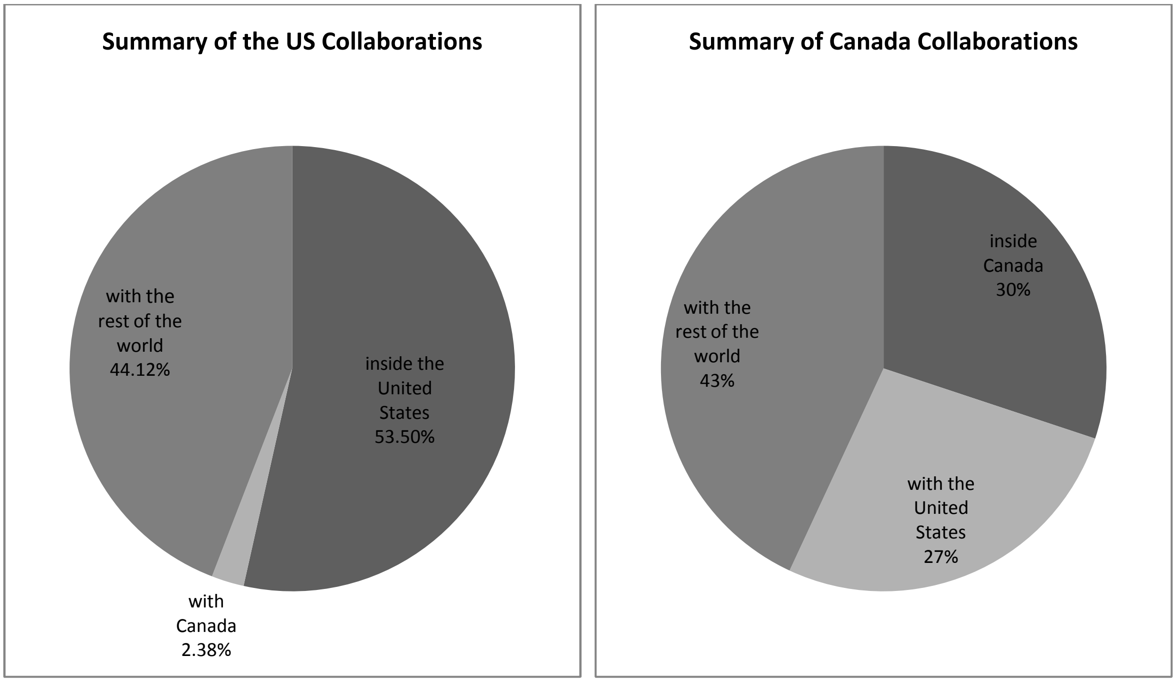


Figure 26 - Summary of average collaboration of Canada and the United States

researcher to find a collaboration partner with an appropriate expertise within the US than it is for a Canadian researcher. These results support our Hypothesis 3.

The study of collaboration patterns in Quebec and the rest of Canada shows an increasing trend in the share of internal collaborations in Quebec. Figure 27 shows the average number of internal and external (rest of Canada) collaborators per each Quebec-based researcher. The increasing trend for internal collaborators and decreasing trend for external collaborators shows that researchers in Quebec have an increasing tendency to form their research partnerships among themselves rather than collaborate with the researchers in the rest of Canada. This finding is also supported by our results presented in Figure 28 where the share of internal and external collaborations of Quebec researchers is illustrated. According to these results, our Hypothesis 4 is accepted.

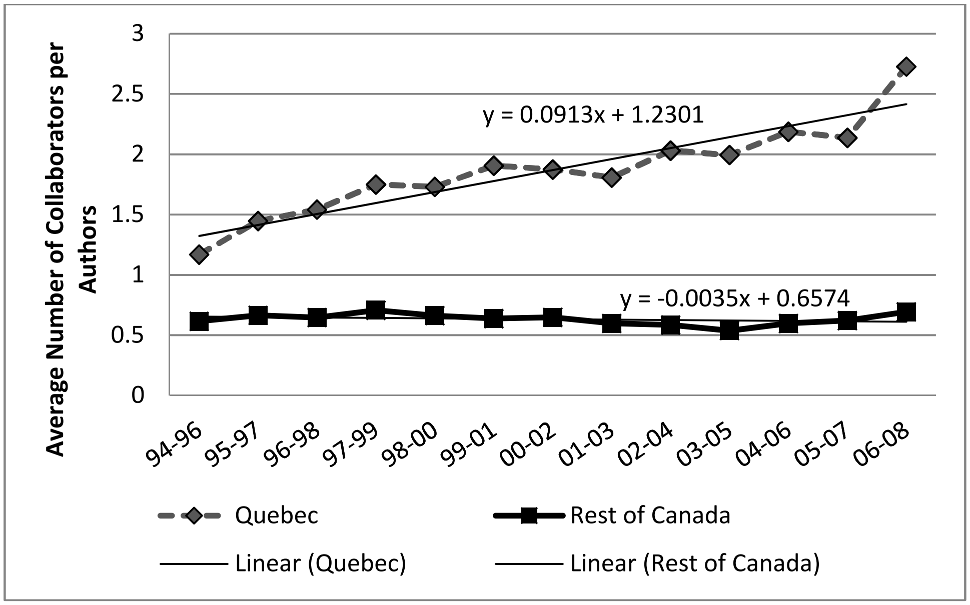


Figure 27- Average number of collaborators per authors for Quebec-based Researchers

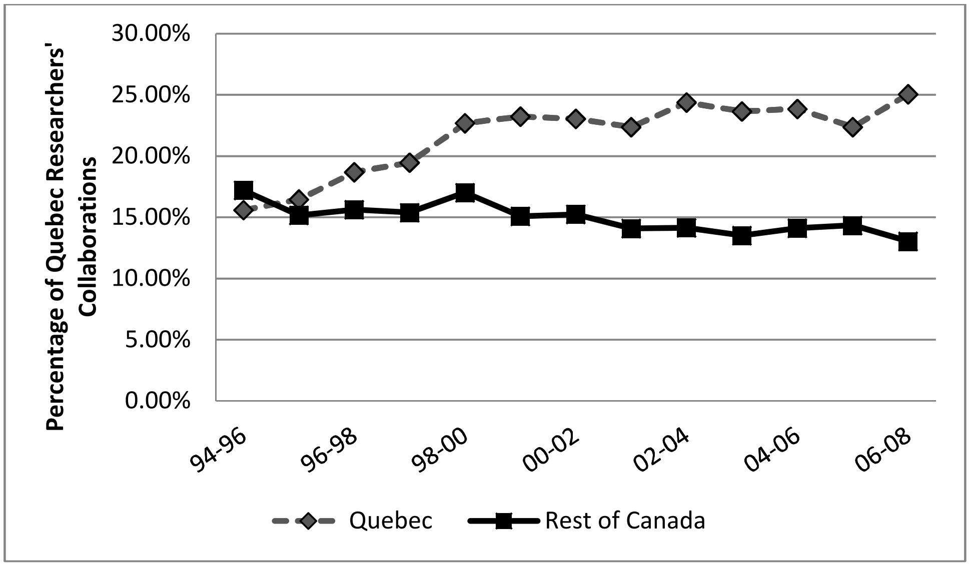


Figure 28 - Percentage of collaboration of Quebec-based researchers inside and outside of Quebec

5.3 Collaboration between Academia and Industry

As explained in previous section, the study of academia and industry collaboration is limited to our article-based networks, because the affiliations of the scientists are not provided in USPTO.

In order to study the collaboration between academia and industry, we split the authors into two groups based on their affiliations. We assume that the authors who are affiliated to universities and schools are academic authors and the rest are considered as non-academic (industry) authors. Non-academic authors include mainly the authors from industry, but also others, for example the authors from governmental labs and research institutes, non-for profit organizations, etc. As explained in 4.3.5, we found the presence of non-industrial affiliations among non-academic sub-category is quite minimal in our database. For authors with more than one affiliation, their first affiliation in Scopus served as the basis for our categorization.

Using this classification, we compare the network characteristics of academic and non-academic subnetworks in three aspects. First, number of collaboration pairs and total number of collaborations within each subnetwork and between them is calculated. Second, the network indicators for centrality and cliquishness for each subnetwork are measured. Finally, we categorized each subnetwork into two smaller subnetworks based on the collaboration with the other group. In other words, we fragmented the network into four subsets; (1) academics who have at least one collaborative relationship with non-academics (hereafter AC-NA), (2) academics with no collaboration with non-academics (hereafter AC-AC), (3) non-academics who have at least one academic collaborator (hereafter NA-AC), and (4) non-academics who have only non-academic collaborators (hereafter NA-NA).

Figure 29 shows the percentage of collaboration pairs in each subnetwork and between them. In other words, this figure shows what percentage of edges of the network link academics between themselves, non-academics with other non-academics, and academic researchers with non-academic ones. Also, Figure 30 shows the share of each group in the total number of collaborations. This means that in the first figure the number of collaborations among two authors is disregarded and only the number of collaborative pairs is considered, whereas in the second figure each instance of collaboration within a collaborating pair is counted. In both figures, we can see that the intra-academic collaborations are most frequent, while the collaborations among non-academics are the least. Also, an increasing trend in the share of purely academic collaborations is observed, while the share of collaboration between two non-academic researchers, and collaboration between academic and non-academic researcher is decreasing. Note that about it is the trends in the percentages which are discussed here, not the numbers. The number of collaboration pairs and the total number of collaborations are increasing in all three types of collaborations.

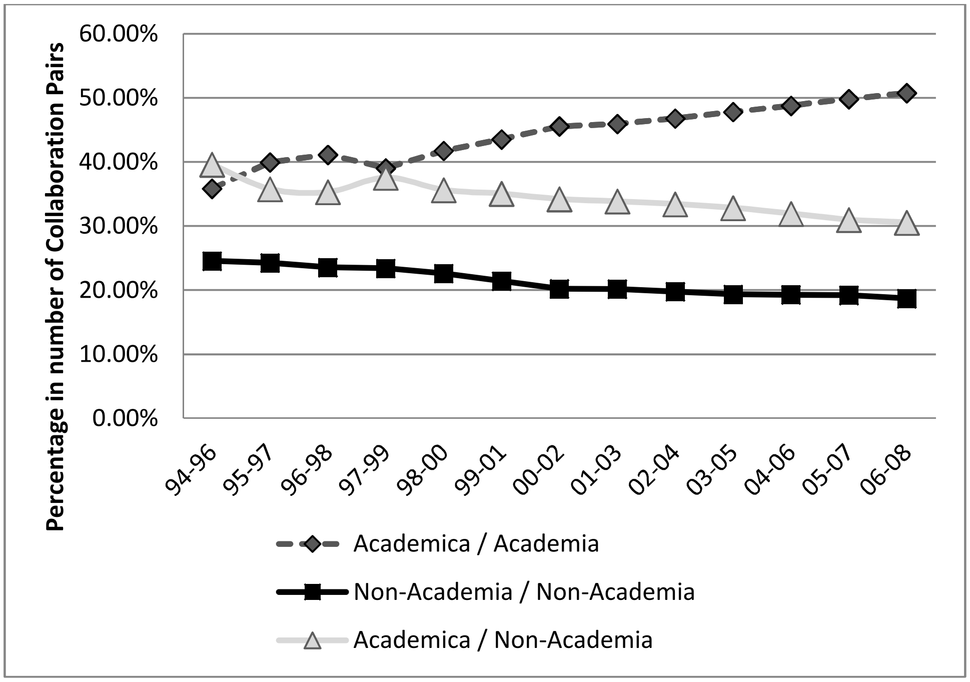


Figure 29 – Percentage of collaboration pairs inside academia and non-academia subnetworks and between them

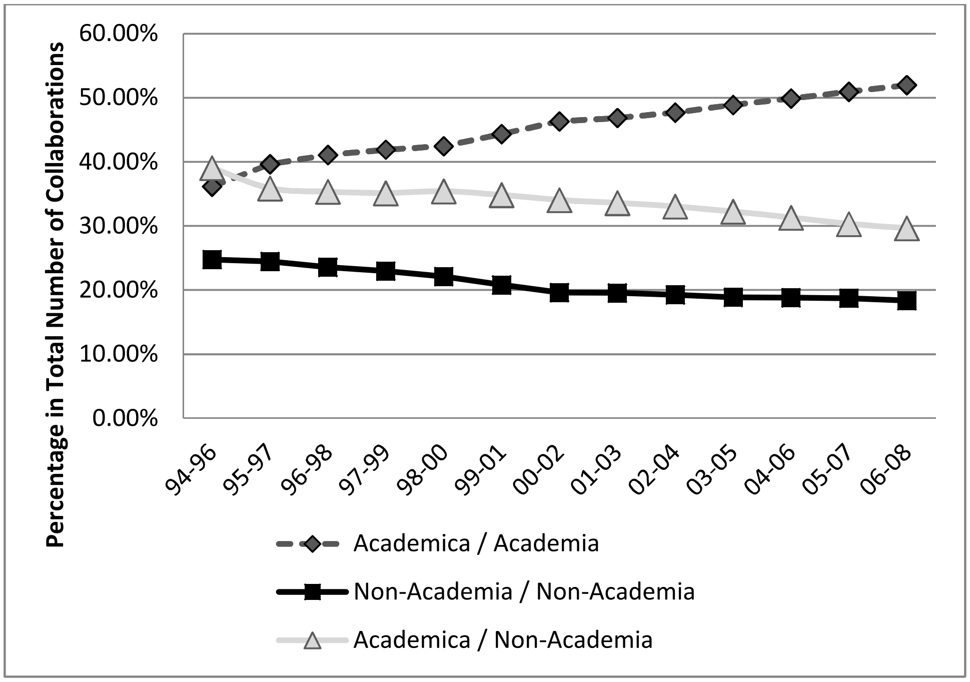


Figure 30 - Percentage of total number of collaboration inside academia and non-academia subnetworks and between them

These results indicate that although the collaborations between the academic and industry worlds account for a smaller share compared to purely academic collaborations, there is still a relatively high percentage of collaborations between the two subgroups (more than one third). This result highlights the important role of collaborations between universities and companies in the field of nanotechnology, and hence it supports our Hypothesis 5.

In order to characterize the structure of these subnetworks and compare the role of authors of each subnetwork, we calculate three network indicators; average clustering coefficient (Figure 31), average betweenness centrality (Figure 32), and average degree centrality (Figure 33).

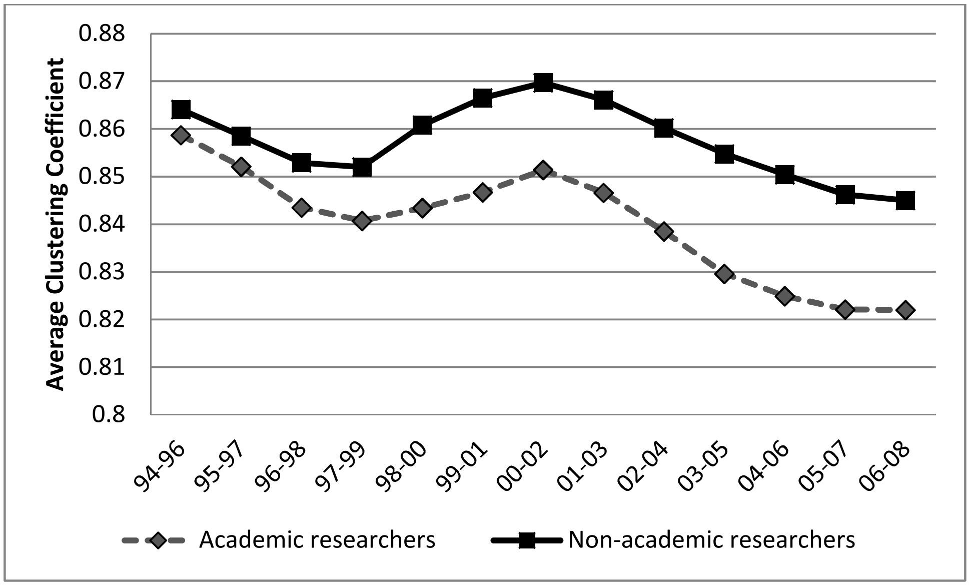


Figure 31- Average clustering coefficient in the network of academics and non-academics

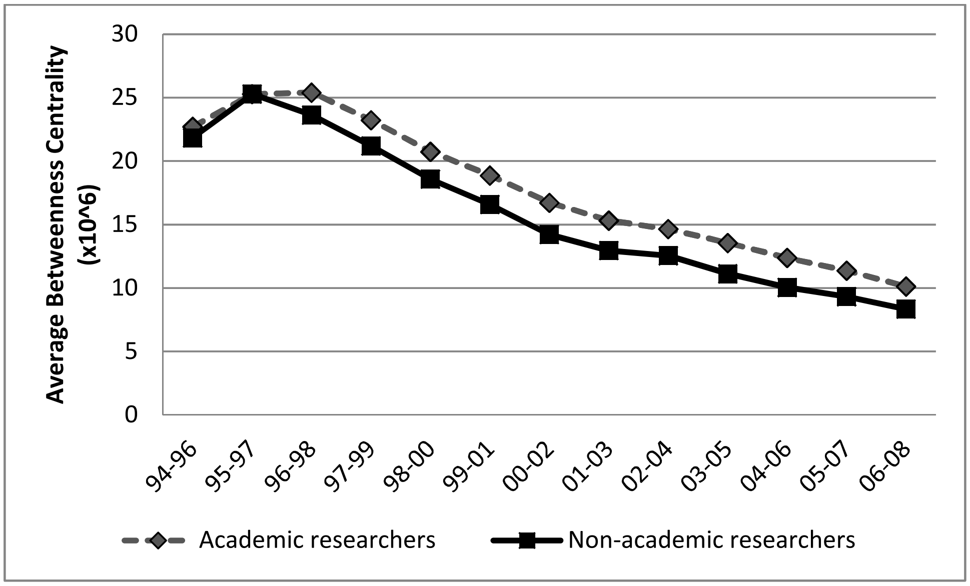


Figure 32 - Average betweenness centrality in the network of academics and non-academics.(values are multiplied by 10^6)

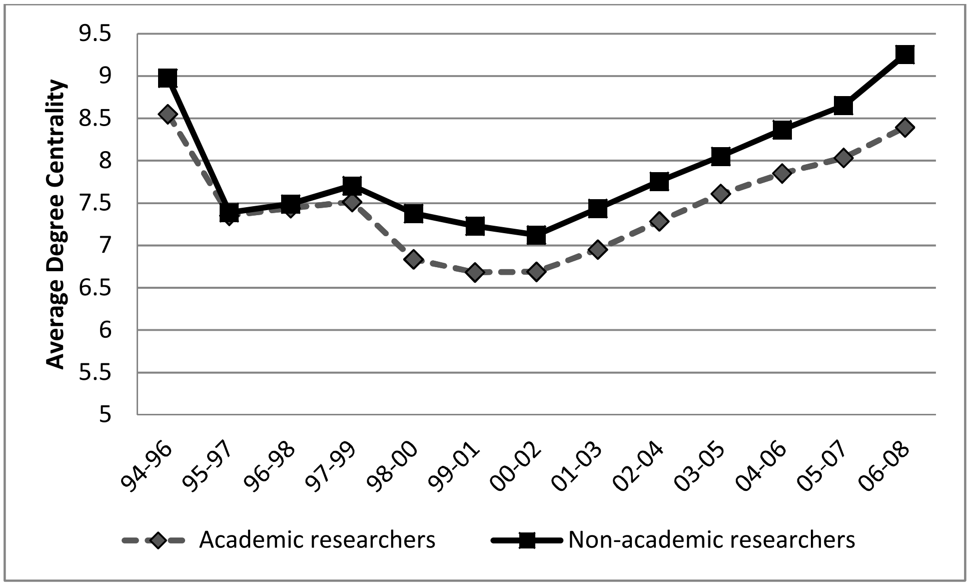


Figure 33 - Average degree centrality in the network of academics and non-academics

The results we obtained do not support our Hypothesis 6. Although we expected a higher cliquishness in academia subnetwork, we found higher values for average clustering coefficient in non-academia subnetwork. Also, higher values of average degree centrality among non-academic researchers show that authors with non-academic affiliations have more collaborators in the network on average. Although the values for average betweenness centrality in the subnetwork of academic researchers are higher than the ones in the non-academic subnetwork, the difference is not significant (note that values are multiplied by 10^6).

The same network indicators are measured for the four subcategories explained before. In this categorization the isolated nodes with no collaboration are ignored. The results of these calculations show that academics and non-academic researchers who create collaborative partnerships only within their own subgroup are more clustered (see Figure 34). On the other hand, the researchers who connect the academia with industry occupy more central positions in terms of both degree centrality (Figure 35) and betweenness centrality (Figure 36). This means researchers who only collaborate inside their community are more clustered, while the average rate of knowledge transfer and number of collaborators are higher for the researchers who links the academia and industry.

Also, based on the results shown in Figure 34 we reject our Hypothesis 7a which proposes that academic researchers who have at least one non-academic collaborator occupy more cliquish positions than the ones who do not collaborate with non-academics. On the other hand, the comparison of average degree centrality and average betweenness centrality shows that academic researchers who have non-academic collaborators occupy more central positions than the ones who do not collaborate with non-academics. These results accept our Hypothesis 7b.

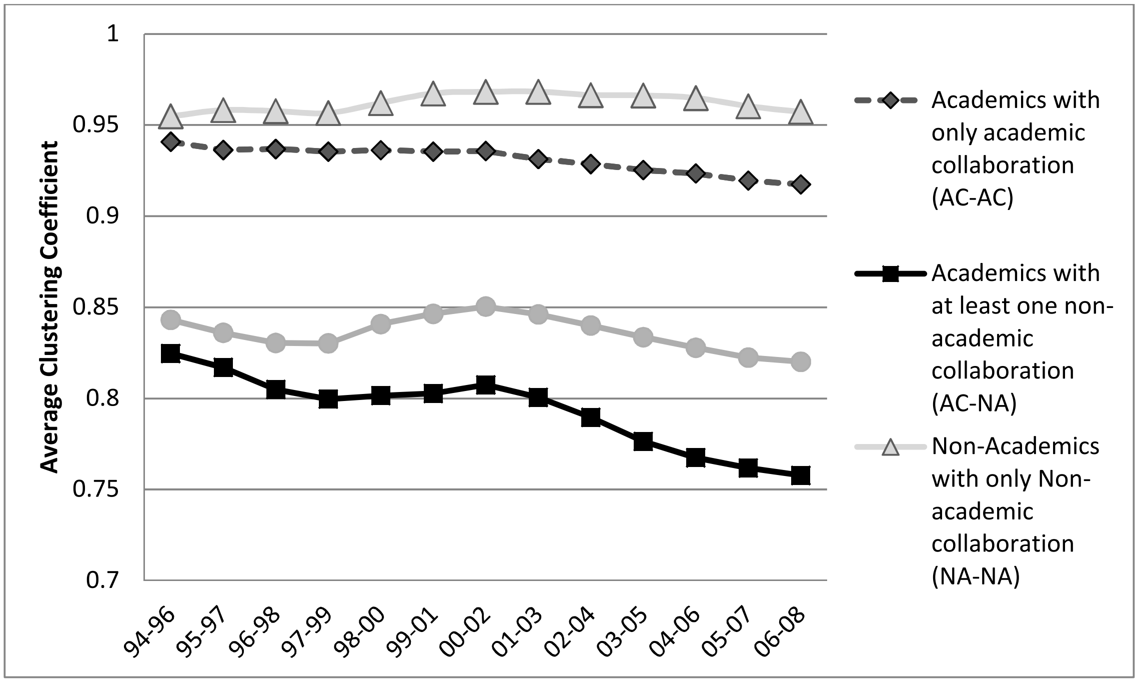


Figure 34 - Average clustering coefficient for academics/non-academics based on their collaboration with the other sub-network

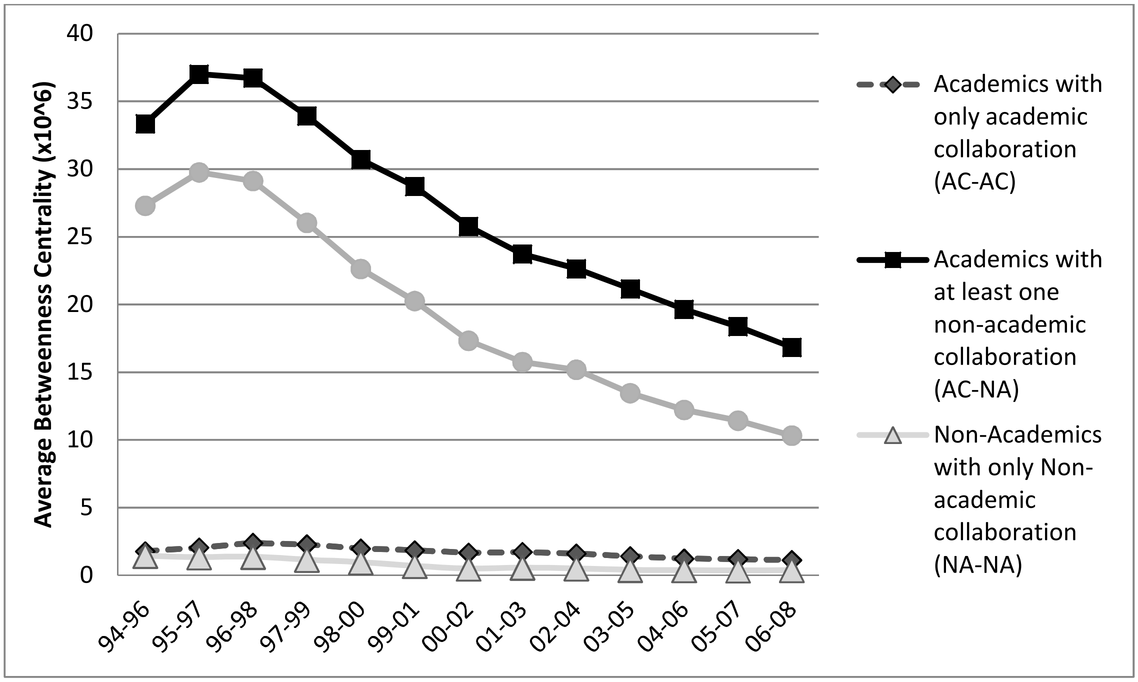


Figure 35 - Average betweenness centrality for academics/non-academics based on their collaboration with the other sub-network (x 10^6)

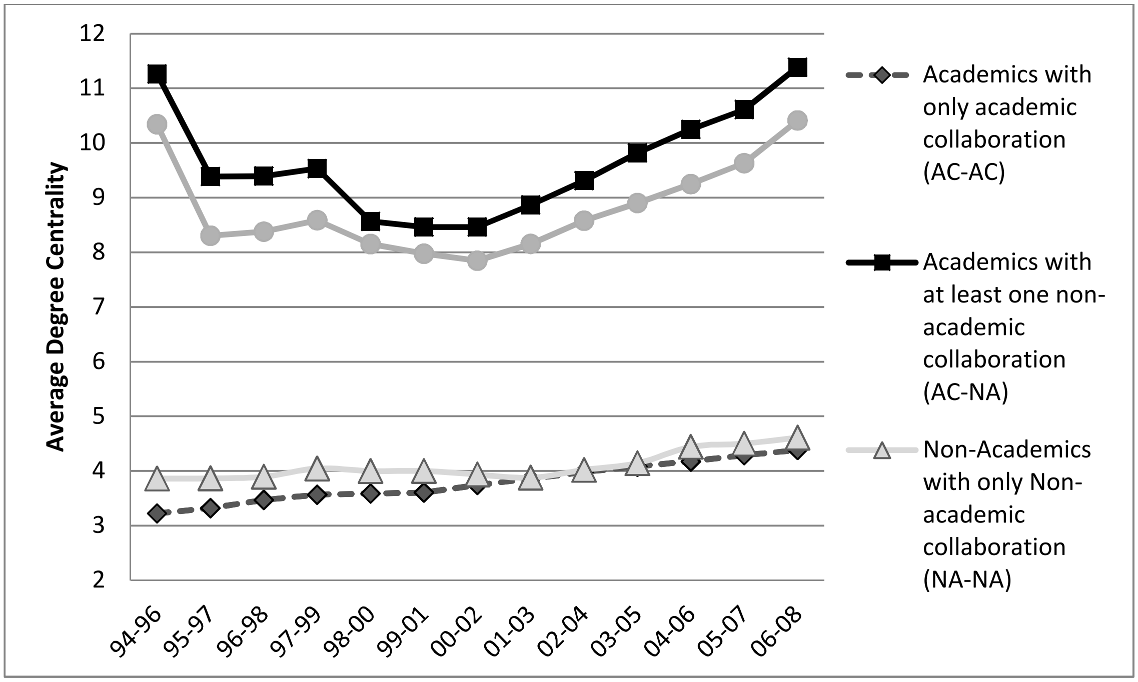


Figure 36 - Average degree centrality for academics/non-academics who collaborates based on their collaboration with the other sub-network

5.4 Regional Differences in the Collaboration between Academia and Industry

In this part of our study, we combine the two previous categorizations based on the residing place and the affiliation of authors. Using this categorization, we compare the collaboration patterns between academia and industry in Quebec, Canada and the United States. The results of these comparisons are presented in two subsections. In the first section, we present and analyze the results comparing Canada with the United States, and then we focus on the Canadian national level, and compare Quebec with the rest of Canada. Results in each section are presented in the following order. First, we compare the share and positions of academic and non-academic researchers in two regions. Then, we categorize the collaborative links into three categories based on the affiliation of collaborators; (1) collaboration between two academic researchers, (2) collaboration between two non-academics, and (3) collaboration between an academic researcher and a non-academic one. Using this grouping in each region, we will find the regional differences in the collaboration between academia and industry.

5.4.1 Academia and Industry in Canada and the United States

According to our results academic nanotechnology scientists are predominant in both Canada and the US. However, the percentage of non-academic nanotechnology researchers is much higher in the US than in Canada. In other words, in both countries the scientific production in nanotechnology is created mainly by the individuals from the academic world, but this trend is much more prevalent in Canada. Figure 37 shows the percentage of non-academic researchers in the network of scientists in Canada and the US. As academics have a greater share in both networks, all the percentages are less than 50%, but there is a significant difference between Americans and Canadians. These results indicate the higher contribution of industry and government in knowledge production in the field of nanotechnology in the United States compared to Canada.

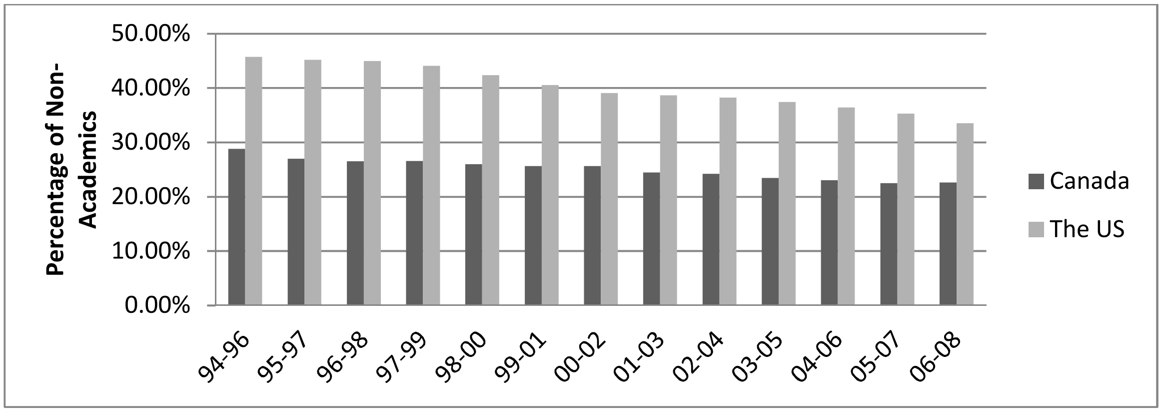


Figure 37 - Comparison of the share of non-academic researchers in nanotechnology network of scientists in Canada and the US

Comparing the academic and non-academic researchers of these two countries is not limited to the number of authors. Also, the position of authors in the world network of nanotechnology scientists is important. In order to compare the position of authors in terms of cliquishness and centrality, we calculate the average of values for clustering coefficient, degree centrality, and betweenness centrality for the authors of each group. According to the results illustrated in Figure 38, there is no significant difference between Canadian and American cliquishness both in academia and non-academia subnetworks. Figure 39 shows the average betweenness centrality in each of the subnetworks and compare Canada with the US. Based on these graphs we can say that, American academic researchers have slightly more central positions than Canadian ones in terms of betweenness centrality. However, the difference in non-academia subnetwork is not significant, especially in the recent years. Finally, Figure 40 illustrates the average degree centrality. In academia, the American authors occupy more central positions in the network than the Canadian researchers in all the periods. However, in non-academic subnetwork there is a difference between 20th and 21st century. In the beginning of 21st century, Americans occupy more central positions than Canadians in non-academic subnetwork, while the opposite results are obtained for the last years of 20th century. This shows the increasing involvement of American non-academic institutions which play more and more important roles in the knowledge production compared to their Canadian counterparts, which accepts our Hypothesis 8a.

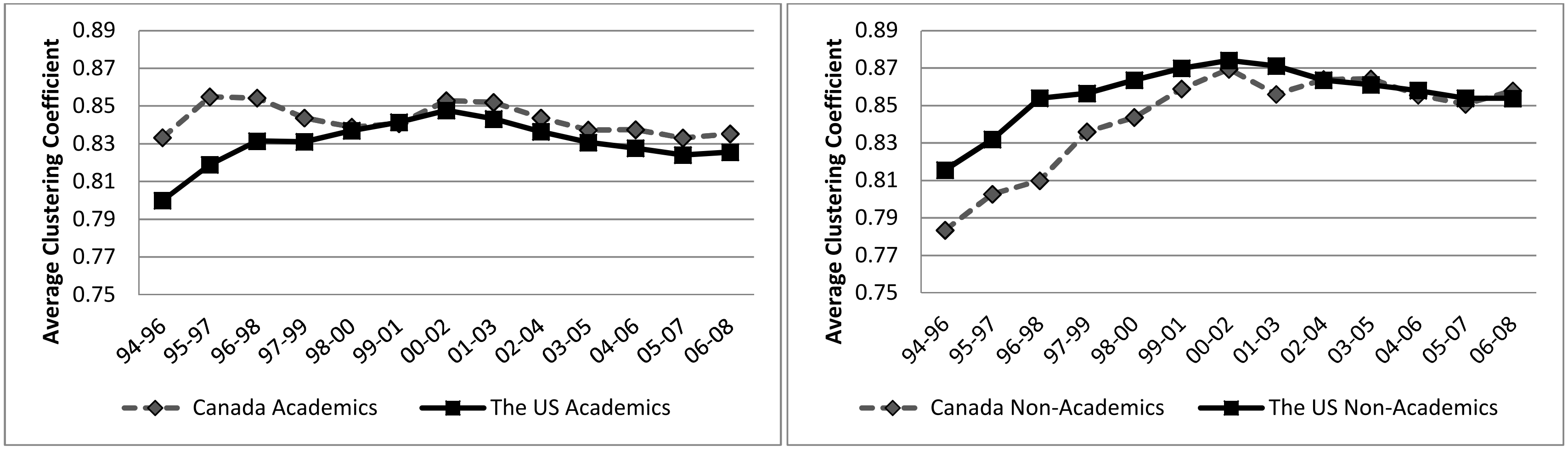


Figure 38 - Average Clustering Coefficient in academics and non-academic communities of Canada and the US; (Left) academic subnetwork; (Right) non-academic subnetwork

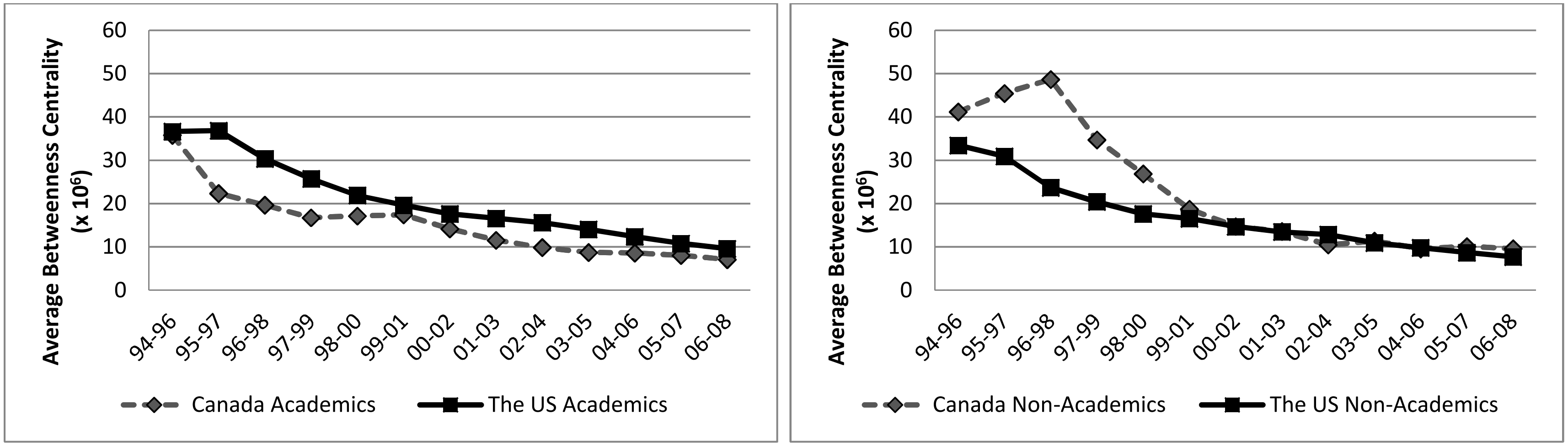


Figure 39 - Average betweenness centrality (x 106) in academics and non-academic communities of Canada and the US; (Left) academic subnetwork; (Right) non-academic subnetwork.

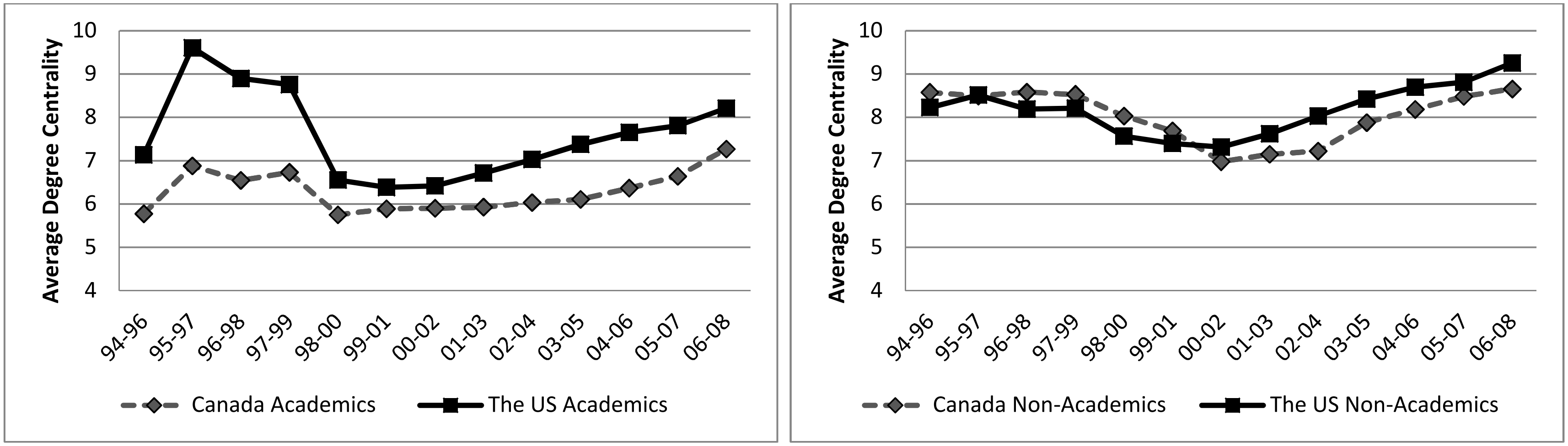


Figure 40 - Average degree centrality in academics and non-academic communities of Canada and the US; (Left) academic subnetwork; (Right) non-academic subnetwork

According to our results comparing academic and non-academic subnetworks of Canada and the US, we can decide about the acceptance and the rejection of some of our hypotheses. Our study shows that American academic scientists occupy more central network positions than the Canadian ones that supports out hypothesis 8b. However, a significant difference in cliquishness is not seen that cannot support or reject our Hypothesis 8c.

Also, based on the results shown in Figure 37, Hypothesis 9a which proposes the greater proportion of non-academics in the US compared to Canada is accepted. However, we cannot accept the hypotheses 9b. We can only claim that American non-academic nanotechnology network tends to be more centralized than Canadian one in terms of degree centrality.

If we divide the authors of scientific articles into academia and non-academia, a higher rate of collaboration between universities and industry is observed in the US compared to Canada (Figure 41a). These results provide support towards accepting our Hypothesis 7a. Also as expected, higher percentage of authors in non-academia results in higher percentage of collaborations between non-academic authors in the US (Figure 41c). Furthermore, an increasing trend is observed in purely academic collaborations both in the US and Canada (Figure 41b), which means the higher activity of universities in each period compared to previous periods. Finally, the average share of all periods is illustrated in Figure 41d and Figure 41e that compare the two regions.

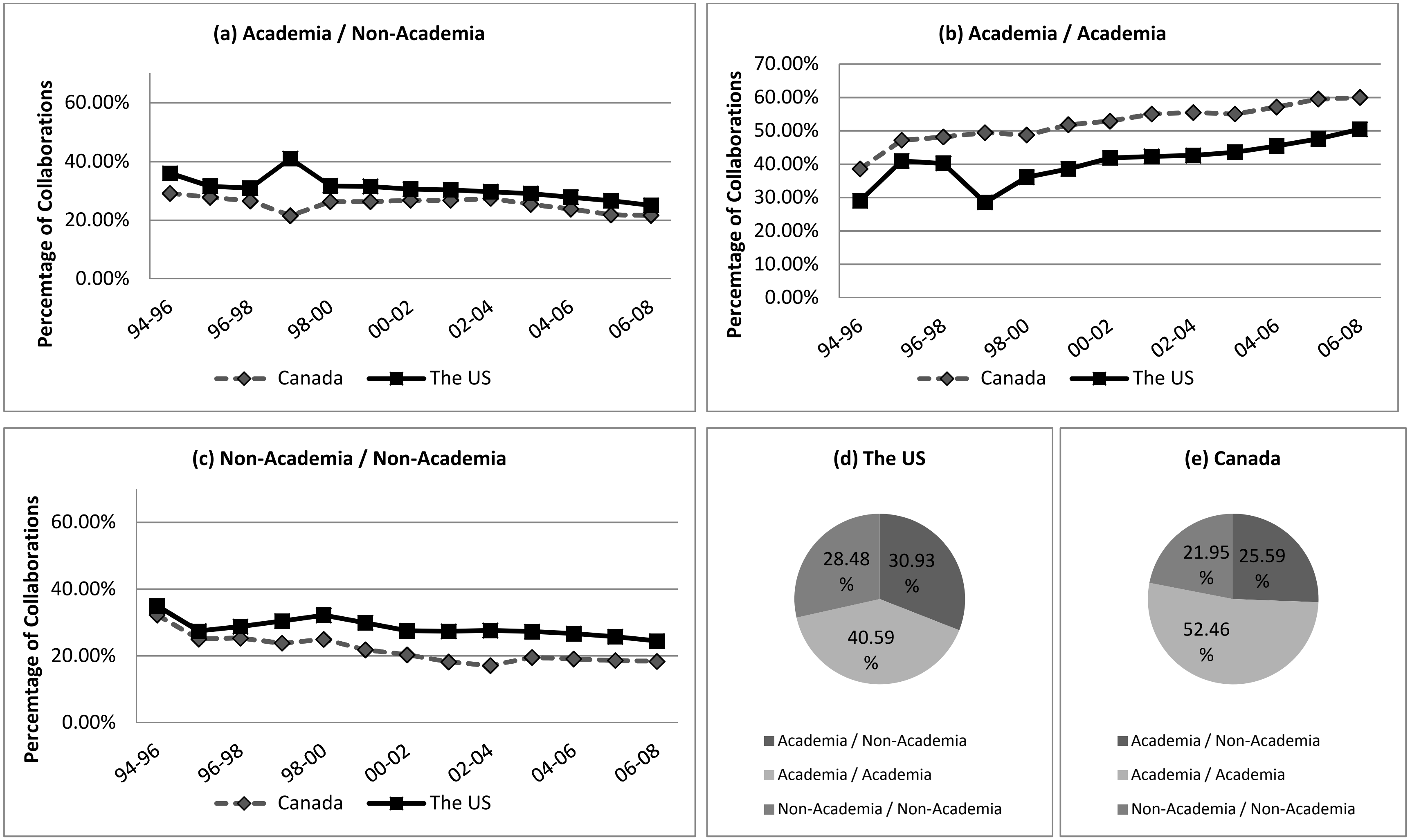


Figure 41 - Comparison of academia and non-academia in the US and Canada based on the collaboration percentages; (a) proportion of collaboration between academia and non-academia in the US and Canada, (b) proportion of collaboration inside academia in the US and Canada, (c) proportion of collaboration inside non-academia in the US and Canada, (d) average share of each type of collaboration in all periods in the US (e) average share of each type of collaboration in all periods in Canada

5.4.2 Academia and Industry in Quebec and the Rest of Canada

Similar indicators are calculated for Quebec and the rest of Canada. According to Figure 42, Non-academic researchers have a smaller share in comparison with the rest of Canada. This indicates that in Quebec, universities are more active in nanotechnology scientific publications than is industry or governmental institutions, in comparison with the rest of Canada.

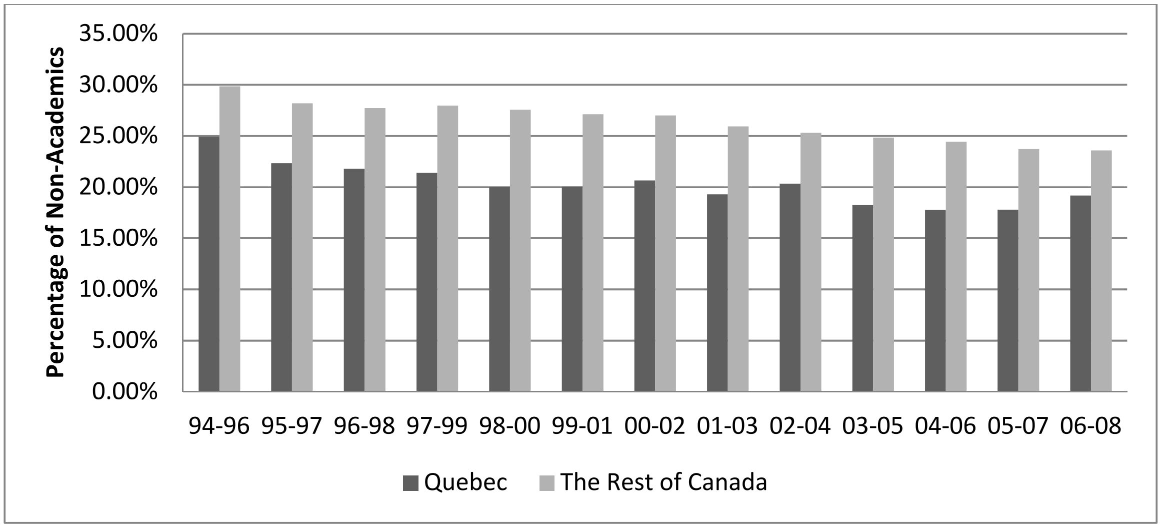


Figure 42 - Comparison of the share of non-academic researchers in nanotechnology network of scientists in Quebec and the rest of Canada

In Figure 43, the average clustering coefficient as an indicator of network cliquishness is measured for both academic and non-academic subnetworks of Quebec and the rest of Canada. As we can see, in academics, the difference is not significant, and the values changing around 0.85 both for Quebecois and other Canadian researchers. However, in non-academics, a significant gap between the results for these two regions is observed. This suggests that, Quebec non-academic researchers are not as clustered as the rest of Canada. The values for average betweenness centrality and average degree centrality are illustrated in Figure 44 and Figure 45. Again the difference between Quebec and the rest of Canada in academia is not significant, and almost follows the pattern of whole Canada. However, the average of centrality values for Quebec is significantly higher than the rest of Canada in non-academic researchers. Therefore, we can claim that, non-academic nanotechnology researchers in Quebec are less clustered, while occupy more central positions compared to the rest of Canada. In other words, although the Quebec-based non-academic researchers do not create relatively dense clusters, they have an important role in the knowledge transfer in the whole network comparing to the researchers of the rest of Canada.

Results of collaborations analysis in academic and non-academic subnetworks of Quebec and the rest of Canada are summarized in Figure 46. Since the share of academic researchers is bigger in Quebec compared to the rest of Canada, a higher share of collaboration in academia and a lower share in non-academia were expected. Figure 46b and Figure 46c are supporting this. Also, Figure 46a, which compares the collaboration between academic and non-academic researchers in these regions, does not show a significant difference. A summary of three types of collaborations for each region is shown in Figure 46d and Figure 46e.

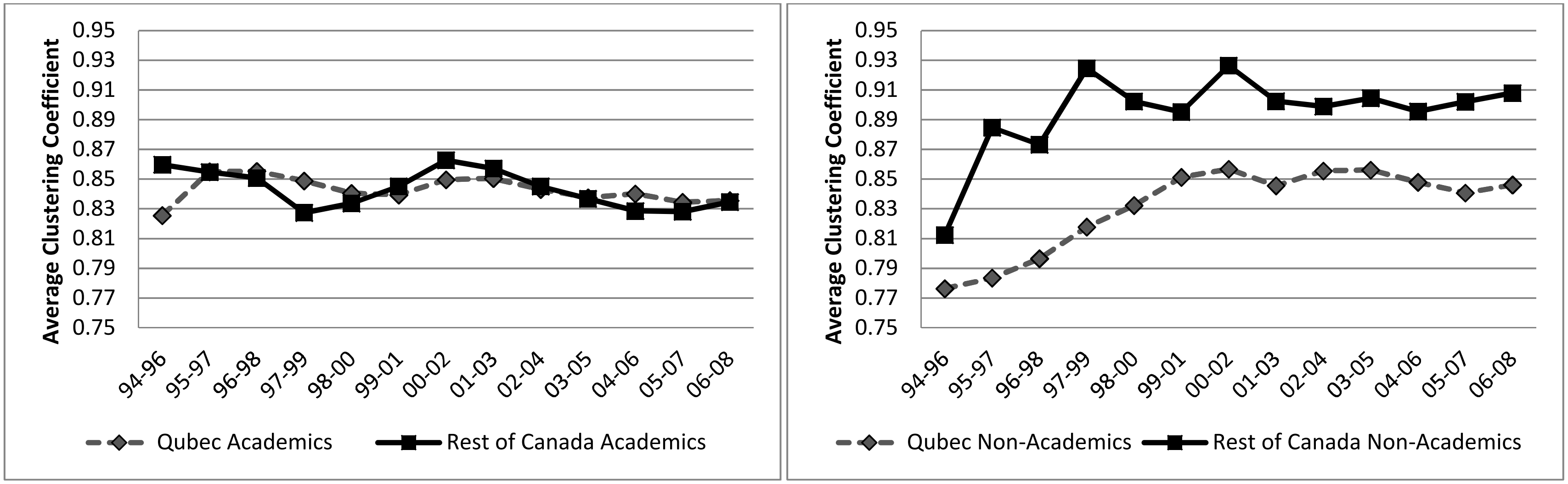


Figure 43- Average clustering coefficient in academics and non-academic communities of Quebec and the rest of Canada; (Left) academic subnetwork; (Right) non-academic subnetwork

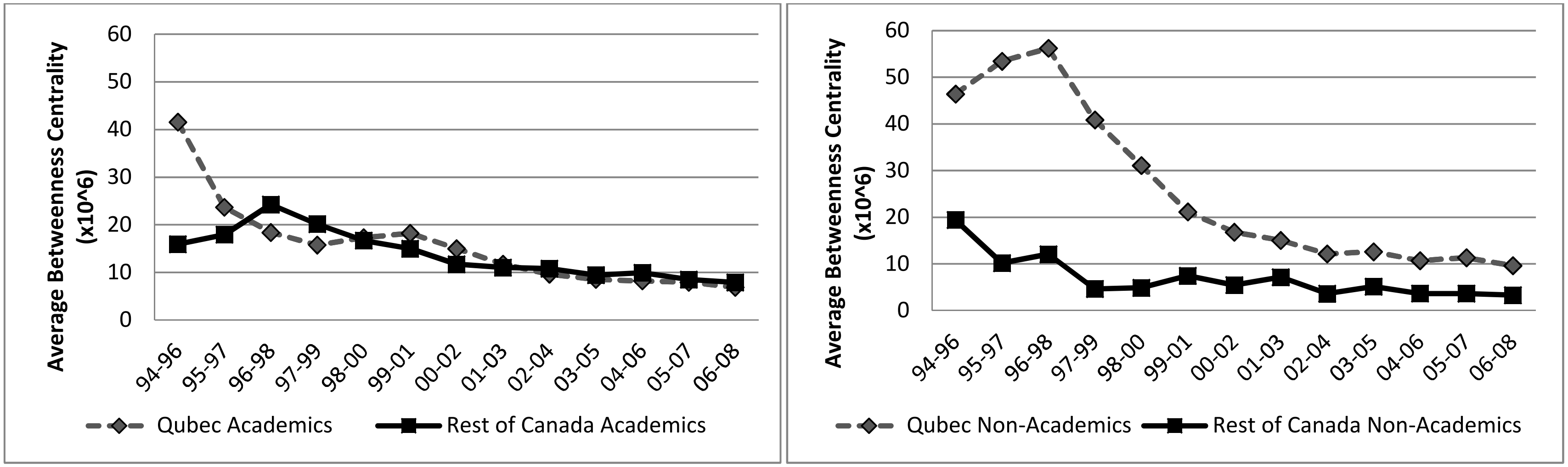


Figure 44 - Average betweenness centrality in academics and non-academic communities of Quebec and the rest of Canada; (Left) academic subnetwork; (Right) non-academic subnetwork

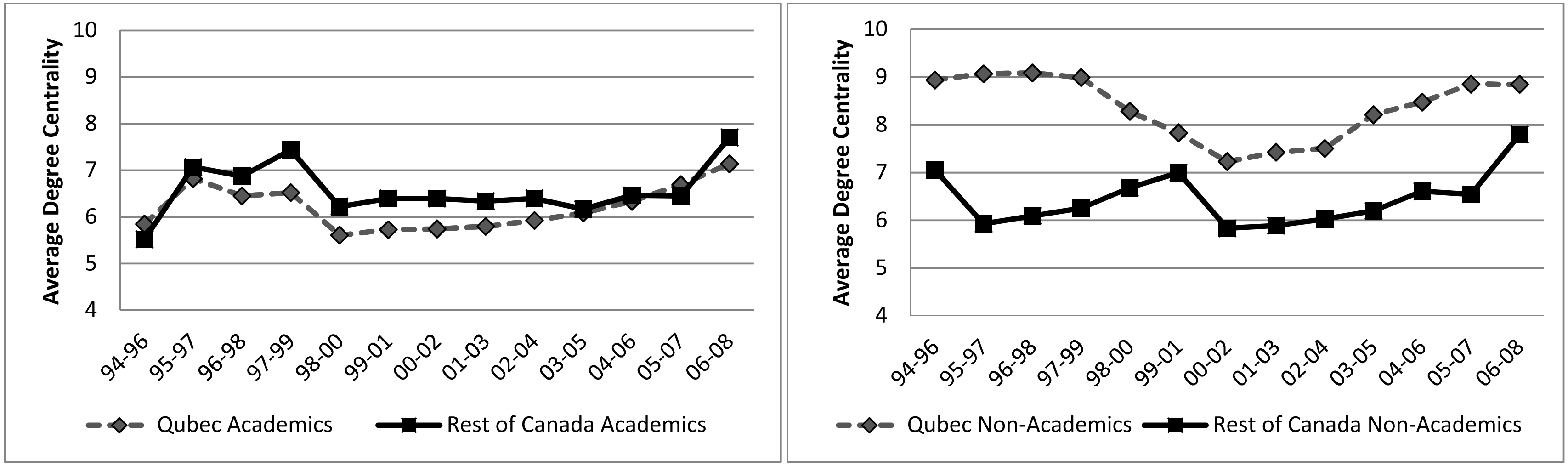


Figure 45 - Average degree centrality (x 10^6) in academics and non-academic communities of Quebec and the rest of Canada; (Left) academic subnetwork; (Right) non-academic subnetwork

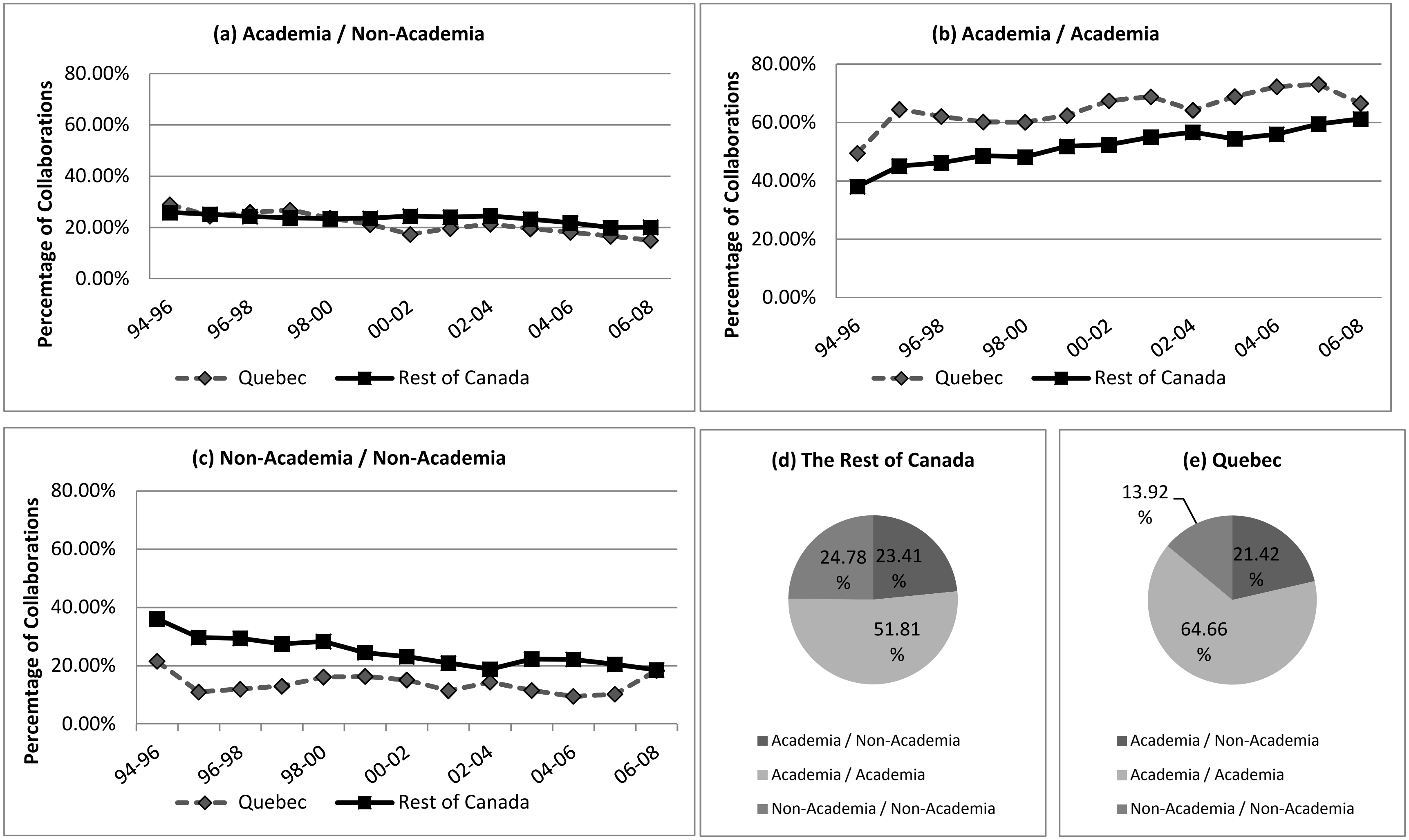


Figure 46 - Comparison of academia and non-academia in Quebec and the rest of Canada; (a) proportion of collaboration between academia and non-academia Quebec and the rest of Canada, (b) proportion of collaboration inside academia in Quebec and the rest of Canada, (c) proportion of collaboration inside non-academia in Quebec and the rest of Canada, (d) average share of each type of collaboration in all periods in the rest of Canada (e) average share of each type of collaboration in all periods in Quebec

6. SUMMARY AND CONCLUSIONS

Nanotechnology is a newly-emerged and fast-growing multidisciplinary field of study that has an influential role in almost every aspect of the economy. Also, this field is known as a platform technology and is predicted to direct the future technological change in the society. Therefore, it is important for nations, like Canada and the United States, to start implementing an infrastructure to increase the performance of facility and innovation management, which is known as national innovation system. National innovation system of each nation is a network of institutions whose interactions determine the performance of the technology and information flow and the knowledge and innovation diffusion. In this study, our focus was on the comparison of innovation transmission and knowledge production in Quebec, Canada, and the United States.

Scientific articles and patents are considered to be a measurable outcome of research and development in science and industry. Many articles and patents are produced through collaboration of several researchers or inventors. Therefore, co-authorship of an article or co-invention of a patent can be a partial indicator of collaboration, assuming that not all the scientific collaborations results in an article and a patent. In order to study the concept of the scientific collaboration, we utilized social networks analysis in this study. In this approach, the community of scientists or inventors is assumed as a social network in which scientists or inventors are the actors, and their collaboration in a research or in an innovative activity (evidenced by an article or a patent) indicates their collaborative connection. Using the indicators introduced by social network scientists, we could characterize the collaboration patterns, and the role of the scientists in the knowledge diffusion and technology transfer in the national innovation systems.

The main purpose of this research was performing a regional comparison of nanotechnology innovation of Quebec, Canada and the United States by studying the characteristics of collaboration networks of scientists and inventors in these regions.

Our comparative analysis involved two phases. We compared the network positions and collaboration patterns of Canadian nanotechnology scientists with the American ones, and then an analogous comparison of the network of Quebec-based scientists with scientists from the rest of Canada was performed. In order to create a more comprehensive picture of nanotechnology innovation and knowledge production two types of networks were built and examined. First, seven "co-invention" networks were created based on the patent collaborations from 1994 to 2002 in three-year intervals. The information of nanotechnology patents and their inventors was extracted from Nanobank database, which is constructed based on the USPTO database. The second type of network is "co-authorship" network. We created three-year co-authorship networks on the basis of scientists' collaboration in the nanotechnology articles from 1994 to 2008. Scopus database was used to extract the nanotechnology scientists and articles.

The first objective of this study was to develop a software program that searches the nanotechnology keywords in the full text of the articles, and extracts the information on the articles and their authors from the Scopus. As Scopus is a database of abstracts and does not offer the full text search, our program gets the results of full text search from Google Scholar, and searches the title of each article in the Scopus. All found articles are stored in a database which we will use for creating the networks. Both the program and the created database were created to be used in further research studies as well. In fact, several other students have already made a benefit from them in their research projects.

The second objective of this research was to characterize the networks by measuring various indicators related to the network positions of each scientist. The place of residence (patents) and the affiliation (articles) were used to assign the inventors and the scientists a region and to divide them into the groups of Quebecois, Canadian and American scientists. We compared the Canadian sub-network with the American one, and also the sub-network of Quebec-based scientists with the scientists from the rest of Canada, in each of the 3-year periods.

Finally, the last objective of this research was to determine and compare the characteristics and the network positions of academic and non-academic scientists within the co-authorship network in Quebec, Canada and the United States. Based on the affiliation of the authors, we divided the authors into two groups: academics and non-academics, then we measured various network properties for each subnetwork separately and compared the findings in each group.

The results of our research are categorized into four sections. First, we studied the fundamental characteristics of each network including size of the network, collaboration patterns, fragmentation, structural cohesion, distance indicators and cliquishness. According to our results, both networks have a high degree of cliquishness (>0.83), and very small values for degree separation, i.e. the average distance of every two reachable scientists (around 6). Also, a decreasing trend in the fragmentation of the networks was observed.

Moreover, we found that the network of scientists does not follow the random model of Erdős and Rényi. On other hand, as we observed the small-world phenomena in the network (relatively low degree of separation coupled with a high degree of cliquishness), we conclude that these networks follow the small-world model of Watts and Strogatz. Finally, we examine the properties of scale-free model of Barabási and Albert in our networks, and as the degree distributions of these networks do not follow the power law, we conclude that these networks are not fitted to scale-free model.

In the second section of results, we present the regional comparisons in both networks. In comparison of collaboration networks in Canada and the United States, we found that, as expected, the American nanotechnology network is bigger, involves more collaboration, and is more centralized than Canadian nanotechnology network. We believe that the higher amount of investment in the field of nanotechnology in the US and the earlier involvement of American scientists in this field are the main reasons for this difference. Also, the collaboration analysis of the Canadian and American scientists shows that in Canada there is a higher tendency of international collaboration (especially with the researchers from the United States), compared to the American scientists who tend to search more for their collaboration partners within their own country. This was also expected as the huge network of nanotechnology in the US provides a more chance for the American scientists to find their collaborators within the US. Given the small population of Canada, the researchers have their choices of finding Canadian experts in specific nanotechnology fields more limited and they need to look for the suitable partnerships outside the country. We also find an increasing share of internal collaboration within Quebec compared to the rest of Canada in co-authorship network. We explain this by the linguistic differences of the Quebec researchers who may be more inclined to opt for a French-speaking collaborator from Quebec instead of a non-French speaker from the other provinces or the United States.

Third section of our results was dedicated to the comparisons of the sub-network of academics with non-academics in the co-authorship network of articles. Our results indicate that the largest amount of collaboration takes place purely within academia, which is followed by collaborating pairs formed by one academic and one non-academic researcher, while the least common collaborative partnerships are among non-academic researchers. An interesting finding was that there is a higher level of cliquishness within non-academic researchers. In other words, although non-academic researchers make fewer collaboration pairs, their collaborations are more clustered, i.e. they create local groups collaborating together. We explain this by the collaboration pattern within nanotechnology companies, which are usually small and medium sized and thus usually focused on a relatively narrow research area. The industrial research teams would also be created mainly within the boundaries of each company, with less chances of collaboration with the competitors. Finally, we categorized the scientists into four categories based on whether they have collaboration with the community that they do not belong to or not. According to our findings, the non-academic scientists who only collaborate in non-academic community are more clustered than non-academics who also collaborate with academics. The same results were obtained for the academics. On the other hand, we found less central position for non-academics and academics that do not collaborate outside their community comparing to the ones who have collaborators in other community. To summarize, we conclude that although the academics (and non-academics) that collaborate only in academia (non-academia) are more clustered, they have a less important position in knowledge transfer between the scientists. This creates an interesting picture of nanotechnology collaboration, where most of the academic scientists and non-academic industrial researchers work within their own community in smaller clusters created around their research specializations, while there are certain connecting agents between the two groups who collaborate with both communities and thus bridge the academic with non-academic world. These researchers occupy central positions in the network and fuel the knowledge transmission on the industry-academia frontier. Contrary to the finding of Balconi et al. [80], we observe that these important individuals are not (almost) strictly coming from academia, but they are created by both academic-based researchers collaborating with non-academicians and the industrial researchers who keep research ties with the university scientists. We explain the different result by the dissimilar networks examined in the two studies. Balconi et al. [80] have investigated the networks which mixed the researchers from various industries together, while we focus on nanotechnology innovation, which is very specific in its organization.

In the last section, the academic and non-academic sub-networks are compared in each pair of regions (Canada and the US, and Quebec and the rest of Canada). According to our results, there is a higher tendency of American nanotechnology academic scientists to collaborate with non-academics compared to their Canadian counterparts. Also, we found that American academic scientists occupy more central and less cliquish network positions than the Canadian ones. This means that Canadian academics form a more clustered sub-network, but they are less involved in transfer of knowledge and innovation in the network. In case of non-academics, we found that Americans have recently started occupying more central positions than Canadian non-academics.

Furthermore, we observe that non-academic researchers are more collaborative in the United States than Canada, though in both regions, academic researcher has a bigger share of collaboration. In the final comparison of this study, we compared Quebec and the rest of Canada in terms of academic and non-academic collaborations. According to our results, Quebec non-academic nanotechnology network is less clustered and accounts for a smaller proportion of the collaborations than the rest of the Canada. However, network positions of the Quebec-based researchers are more central, which prove the higher impact of Quebec non-academic network in the knowledge transfer than non-academic network of the rest of Canada. This result was somehow unexpected, as we previously found a great amount of Quebec-based internal collaborations, which we explained by the linguistic characteristics of the Quebec region. Here, however we observe that this is not valid within the non-academic world. Our results here show that the Quebec industrial researchers are in fact more open and working with the Anglophone world more than the Quebec academicians, and thus serve as important agents of knowledge transmission. Moreover, here we also observe that Quebec academic researchers have proportionally much more intra-academic collaborations than the rest of Canada. If we put all these results together we can clearly discern a distinct pattern of Quebec nanotechnology innovation, where much of the nanotechnology knowledge is created through relatively closed internal collaborative partnerships of Quebec academia. The university researchers however need to rely on the non-academic researchers which are geographically better interconnected and thus have a better access to the knowledge circulating within the international collaboration network in order to bring to Quebec fresh ideas originating in distant places. This shows important but diverse roles played by academic and industrial researchers in the Quebec nanotechnology knowledge production.

In a nutshell, the comparative analysis performed in this research between three regions (Quebec, Canada, and the US), and two communities (academic and non-academic) provides the differences and similarities in the trend of improvement in the nanotechnology in each of the subgroups.

7. FUTURE WORKS AND REMARKS

There are many avenues for future research. Here are some examples of how this study can be extended in order to further explore the topic:

In this study, we used the number of collaborations, publications and patents to compare the regions. All these factors show the quantity of the knowledge diffusion and innovation transfer in this field, not the quality. In other words, we assume all the publications and patents have the same effect in the nanotechnology and nanoscience, and we ignore the differences in the level of collaborations. Other studies can develop a framework in which the quality of articles and patents, and also the level of collaborations can be studied. In this framework, the weight of the network edges can be assigned based on both quality and quantity of the collaborations.

Also, the comparison of Quebec, Canada and the United States can be done at the province or state level or even at the level of regional clusters. With the aid of this comparison, the pioneer regions in the field of nanotechnology in terms of quantity and quality of collaborations and also the critical positions in the network can be identified.

Moreover, other studies can extend the results of our research into the global level. Although we did not ignore the role of other countries in calculating the network indicators, our focus was on the mentioned regions in the North America.

Finally, some limitations that we had in the databases can be resolved for the future studies. For example, the limitation in the patent dataset of Nanobank restricts our result up to 2002. A more comprehensive database of patents can strengthen the results of this study. Also, using the Scopus database limits us the English articles, while articles in other languages like French, Spanish or Chinese have a significant share in the knowledge diffusion. Other studies can merge the data from the databases in the other languages with the current database to include a wider range of articles.

BIBLIOGRAPHY

[1] M. E. Ludeña, J. Pérez A, E. Fernández T, A. Fischmann and G. A. Plonski, "Assessing Innovation Networks in Nanotechnology," in *IV Globelics Conference*, Mexico City, 2008.

[2] T. Heinze, "Nanoscience and Nanotechnology in Europe: Analysis of Publications and Patent Applications including Comparisons with the United States," *Nanotechnology Law & Business,* vol. 1, no. 4, p. 2–20, 2004.

[3] M. R. Darby and L. G. Zucker, "Grilichesian Breakthroughs: Invention of Methods of Inventing and Firm Entry in Nanotechnology," *NBER Working Paper Series,* p. Working Paper No. 9825, 2003.

[4] K. Miyazaki and N. Islam, "Nanotechnology systems of innovation—An analysis of industry and academia research activities," *Technovation,* vol. 27, p. 661–675, 2007.

[5] S. J. H. Graham and M. Iacopetta, "Nanotechnology and the Emergence of a General Purpose Technology," http://ssrn.com/abstract=1334376, 2009.

[6] J. Canton, "Global Future Business. The Strategic Impact of Nanoscience on the Future of Business and Economics," Institute for Global Futures, 1999.

[7] P. Shapira, J. Youtie and L. Kay, "National innovation systems and the globalization of nanotechnology innovation," *Journal of Technology Transfer,* vol. First Online, 2011.

[8] R. Freeman and K. Shukla, "Jobs in Nanotech – Creating a Measure of Job Growth," National Bureau of Economic Research, 2008.

[9] K. E. Drexler, Engines of Creation 2.0: The Coming Era of Nanotechnology - Updated and Expanded, WOWIO Books, 2007.

[10] T. Heinze, "Emergence of Nano S&T in Germany. Network Formation and Company Performance," *Fraunhofer ISI Discussion Papers Innovation System and Policy Analysis,* 2006.

[11] M. Roco and W. Bainbridge, "Societal implications of nanoscience and nanotechnology: Maximizing human benefit," 2001.

[12] J. Canton, "Finding Prosperity: The Innovation Economy," in *The extreme future*, New York, Penguin Group, 2006.

[13] M. D. Mehta, "Nanoscience and Nanotechnology: Assessing the Nature of Innovation in These Fields," *Bulletin of Science, Technology & Society,* vol. 22, no. 4, pp. 269-273, 2002.

[14] M. F. Yegul, M. Yavuz and P. Guild, "Nanotechnology: Canada's Position in Scientific Publications and Patents," in *PICMET 2008 Proceedings*, Cape Town, 2008.

[15] "Canadian Workshop on Multidisciplinary Research on Nanotechnology: Gaps, Opportunities and Priorities," Canadian Institutes of Health Research, Edmonton, Alberta, 2008.

[16] A. Schiffauerova and C. Beaudry, "Is Canadian intellectual property leaving Canada? The study of nanotechnology patenting," *Journal of Technology Transfer,* vol. 36, no. 6, pp. 665-679, 2011.

[17] B. Godin, "National Innovation System; The System Approach in Historical Perspective," *Science, Technology, & Human Values,* vol. 34, no. 4, pp. 476-501, 2009.

[18] C. Freeman, Technology and Economic Performance: Lessons from Japan, London: Pinter, 1987.

[19] R. Nelson, National Innovation Systems. A Comparative Analysis, New York/Oxford: Oxford University Press, 1993.

[20] P. Patel and K. Pavitt, "The Nature and Economic Importance of National Innovation Systems," *STI Review,* vol. 3, no. 14, pp. 9-32, 1994.

[21] S. Metcalfe, "The Economic Foundations of Technology Policy: Equilibrium and Evolutionary Perspectives," in *Handbook of the Economics of Innovation and Technological Change*, Oxford (UK)/Cambridge (US), Blackwell Publishers, 1995.

[22] B. A. Lundvall, National Innovation Systems: Towards a Theory of Innovation and Interactive Learning, London: Pinter, 1992.

[23] OECD, National Innovation Systems, Paris: Organization for Economic Cooperation and Development, 1997.

[24] C. Edquist, "Systems of Innovation; Prespectives and Challenges," in *The Oxford Handbook of Innovation*, New York, Oxford University Press, 2005.

[25] J. Niosi, Canada's national system of innovation, Montreal: McGill-Queen’s University, 2000.

[26] OCED, Managing National Innovation Systems, Paris: Organization for Economic Cooperation and Development, 1999.

[27] D. Mcfetridge, "National Innovation System: A Comparative Analysis," in *The Canadian System of Industrial Innovation*, New York, Oxford University Press, Inc., 1993, pp. 299-321.

[28] G. Stanely, "Upgrading Canada's national innovation system: more than money required," *IRPP Policy Options,* Vols. July-Aug , pp. 68-74, 2007.

[29] "Québec Policy on science, technology and innovation," Conseil de la science et de la technologie du Québec, Québec, 2002.

[30] S. Holtz, "Discussion Paper on a Policy Framework for Nanotechnology," Canadian Institute for Environmental Law and Policy , Toronto, 2007.

[31] "Nanotechnology: a quickly emerging field promising benefits and significant risks," Canadian Institute for Environmental Law and Policy , Toronto, 2008.

[32] J. Pelley and M. Saner, "International Approaches to the Regulatory Governance of Nanotechnology," Regulatory Governance Initiative, Ottawa, 2009.

[33] K. Kuroiwa, "Nanotechnology in Canada," *Japan Nanonet Bulletin,* no. 76, 2006.

[34] K. L. Simons and J. L. Walls, "The U.S. National Innovation System," in *Encyclopedia of Technology and Innovation*, Wiley-Blackwell, 2008.

[35] K. Hill, Univerisities in the U.S. National Innovation System, Tempe, Arizona:
L. William Seidman Research Institute, W. P. Carey School of Business,
Arizona State University, 2006.

[36] D. Mowery and N. Rosenberg, "The U.S. National Innovation System," in *National Systems of Innovation: A Comparative Analysis*, New York, Oxford University Press, 1993.

[37] D. C. Mowery, "Nanotechnology and the US national innovation system: continuity and change," *Springer,* 2011.

[38] A. Hullmann, "The economic development of nanotechnology - An indicators based analysis," Directorate-General for Research, Unit Nano S&T- Convergent

Science and Technologies, 2006.

[39] H. Chen and M. C. Roco, "Worldwide nanotechnology literature publication trend using the Thomson SCI Database," in *Mapping Nanotechnology Innovations and Knowledge: Global and Longitudinal Patent and Literature Analysis*, New York, Springer, 2008.

[40] W. Glanzel, M. Meyer, M. Plessis, B. Thijs, T. Magerman, B. Schlemmer, K. Debackere and R. Veugelers, "Nanotechnology, analysis of an emerging domain of scientific and technological endeavor," Steunpunt O&O Statistieken , Leuven, Belgium, 2003.

[41] J. Youtie, P. Shapira and A. L. Porter, "Nanotechnology publications and citations by leading countries and blocs," *Journal of Nanoparticle Research,* vol. 10, no. 6, pp. 981-986, 2008.

[42] M. C. Roco, "The long view of nanotechnology development: the National Nanotechnology Initiative at 10 years," *J Nanopart Res,* vol. 13, p. 427–445, 2011.

[43] Y. Dang, Y. Zhang, L. Fan, H. Chen and M. C. Roco, "Trends in worldwide nanotechnology patent applications: 1991 to 2008," *J Nano part Res.,* vol. 12, pp. 687-706, 2010.

[44] M. E. J. Newman, "The structure and function of complex networks," *SIAM Rev.,* vol. 45, no. 2, pp. 167-256, 2003.

[45] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D. Hwang, "Complex networks: Structure and dynamics," *Physics Reports,* vol. 424, p. 175 – 308, 2006.

[46] A.-L. Barabási, Linked: The New Science of Networks, Cambridge, MA: Perseus, 2002.

[47] W. d. Nooy, A. Mrvar and V. Batagelj, Exploratory Social Network Analysis with Pajek, Cambridge: Cambridge University Press, 2005.

[48] A. Arenas, A. Diaz-Guilera, J. Kurths, Y. Moreno and C. Zhou, "Synchronization in complex networks," *Physics Reports,* vol. 469, pp. 93--153, 2008.

[49] J. P. Scott., Social Network Analysis: A Handbook, 2 ed., Sage Publications, 2000..

[50] M. E. J. Newman and J. Park., "Why social networks are different from other types of networks," *Physical Review E,* vol. 68, 2003.

[51] D. Easley and J. Kleinberg, Networks, Crowds, and Markets: Reasoning About a Highly Connected World, Cambridge Univ Press, 2010.

[52] A. Kirman, "The economy as an evolving network," *Journal of Evolutionary Economics,* vol. 7, pp. 339-353, 1997.

[53] A.-L. Barabási and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nature Reviews Genetics,* vol. 5, pp. 101-113, 2004.

[54] Z. N. Oltvai, A.-L. Barabasi, S. P. Mason and H. Jeong, "Lethality and centrality in protein networks," *Nature,* vol. 411, pp. 41-42, 2001.

[55] M. E. J. Newman, "The structure of scientific collaboration networks," in *PNAS*, USA, 2001.

[56] A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert and T. Vicsek, "Evolution of the social network of scientific collaborations," *Physica A: Statistical Mechanics and its Applications,* pp. 590-614, 2002.

[57] L. d. F. Costa, F. A. Rodrigues, G. Travieso and P. R. Villas Boas, "Characterization of Complex Networks: A Survey of measurements," *Advances in Physics,* vol. 56, no.

1, pp. 167-242, 2007.

[58] S. Milgram, "The small world problem," *Psychology today,* vol. 1, no. 1, pp. 60-67, 1967.

[59] D. J. Watts and S. H. Strogatz, "Collective dynamics of `small-world' networks," *Nature,* vol. 393, pp. 440-442, 1998.

[60] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social network,* vol. 1, pp. 215-239, 1979.

[61] D. B. West, Introduction to Graph Theory, 2nd Ed, New Jersy: Pearson Education Inc., 2001.

[62] U. Brandes, "A Faster Algorithm for Betweenness Centrality," *Journal of Mathematical Sociology,* vol. 25, p. 163–177, 2001.

[63] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry,* vol. 40, pp. 25-41, 1977.

[64] J. M. Anthonisse, "The rush in a directed graph," Stichting Mathematisch Centrum, Amesterdam, 1971.

[65] D. Koschützki, K. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl and O. Zlotowski, "Centrality Indices," in *Lecture Notes in Computer Science: Network Analysis*, Berlin / Heidelberg, Springer, 2005, pp. 16-61.

[66] G. Sabidussi, "The centrality index of a graph," *Psychometrika,* vol. 31, pp. 581-603, 1966.

[67] C. Beaudry and A. Schiffauerova, "Biotechnology and Nanotechnology Innovation Networks in Canadian Clusters," in *Innovation Networks and Clusters*, Brussels,

P.I.E Peter Lang, 2010, pp. 159-197.

[68] M. E. J. Newman, Networks, an introduction, Oxford Press, 2010.

[69] C. Goffman, "And what is your Erdös number?," *The American Mathematical Monthly,* vol. 76, p. 791, 1969.

[70] R. De Castro and J. W. Grossman, "Famous trails to Paul Erdős," *Mathematical Intelligencer,* vol. 21, no. 3, pp. 51-53, 1999.

[71] M. E. J. Newman, A.-L. Barabási and D. J. Watts, The structure and dynamics of networks, New Jersey: Princeton University Press, 2006.

[72] F. J. Acedo, C. Barroso, C. Casanueva and J. L. Galán, "Co-Authorship in Management and Organizational Studies: An Empirical and Network Analysis," *Journal of Management Studies,* pp. 957-983, 2006.

[73] M. E. J. Newman, "Scientific collaboration networks. I. Network construction and fundamental results," *Physical Review,* 2001.

[74] M. E. J. Newman, "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality," *Physical Review,* 2001.

[75] J. Moody, "The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999," *American Sociological Review,* vol. 69, no. 2, pp. 213-238, 2004.

[76] W. Glänzel and A. Schubert, "Analyzing Scientific Networks Through Co-authorship," in *Handbook of Quantitative Science and Technology Research*, Netherlands, Kluwer Academic Publishers, 2004, pp. 257-276.

[77] C. Beaudry and M. Clerk-lamalice, "Grants, Contracts, and Networks: What Influences Biotechnology Scientific Production," in *Opening Up Innovation:*

Strategy, Organization and Technology, London, 2010.

[78] S. Breschi and F. Lissoni, "Mobility and social networks: localized knowledge spillovres revisited," *CESPRI Working Papers,* vol. 142, 2003.

[79] S. Breschi and F. Lissoni, "Knowledge networks from patent data: Methodological issues and research targets," *CESPRI Working Papers,* vol. 150, 2004.

[80] M. Balconi, S. Breschi and F. Lissoni, "Networks of inventors and the role of academia: an exploration of Italian patent data," *Research Policy,* p. 127–145, 2004.

[81] U. Cantner and H. Graf, "The network of innovators in Jena: An application of social network analysis," *Research Policy,* vol. 35, pp. 463-480, 2006.

[82] J. Singh, "Collaborative networks as determinants of knowledge diffusion," *Management Science,* vol. 51, no. 5, pp. 756-770, 2005.

[83] S. Gauvin, "Networks of Innovators: Evidence from Canadian Patents," *Group Decision and Negotiation,* pp. 411-428, 1995.

[84] C. Beaudry and A. Schiffauerova, "Impacts of Collaboration and Network Indicators on Patent Quality: The Case of Canadian Nanotechnology Innovation," *European Management Journal,* 2011.

[85] J. Guan and Y. Shi, "Small world and collaboration network: Evidence from China's nanotech patenting," *Journal of Informetrics,* 2010.

[86] L. Fleming, C. King III and A. Juda, "Small Worlds and Regional Innovation," *Harvard Business School,* 2006.

[87] A. Schiffauerova and C. Beaudry, "Canadian Nanotechnology Innovation Networks: Intra-cluster, Inter-cluster and Foreign Collaboration," *Journal of Innovation*

Economics, vol. 2, no. 4, pp. 119-146, 2009.

[88] P. Dasgupta and P. David, "Toward a new economics of science," *Research Policy,* vol. 23, pp. 487-521, 1994.

[89] F. Murray, "Innovation as co-evolution of scientific and technological networks: Exploring tissue engineering," *Research Policy,* vol. 31, pp. 1389-1403, 2002.

[90] L. Zucker, M. Darby and J. Armstrong, "Geographically localized knowledge: Spillovers or markets?," *Economic Inquiry,* vol. 36, no. 1, pp. 65-86, 1998.

[91] S. Breschi and C. Catalini, "Tracing the links between science and technology: An exploratory analysis of scientists' and inventors' networks," *Research Policy ,* vol. 39, no. 1, pp. 14-26, 2010.

[92] M. S. Meyer, "Patent citation analysis in a novel field of technology: An exploration of nano-science and nano-technology," *Scientometrics,* vol. 51, no. 1, pp. 163-183, 2001.

[93] E. Noyons, R. Buter, A. van Raan, U. Schmoch, T. Heinze, S. Hinze and R. Rangnow, "Mapping Excellence in Science and Technology across Europe Nanoscience and Nanotechnology," European Commission, 2003.

[94] C. Warris, "Nanotechnology Benchmarking Project," Australian Academy of Science, 2004.

[95] A. L. Porter, J. Youtie, P. Youtie and D. J. Schoeneck, "Refining search terms for nanotechnology," *Journal of Nanoparticle Research,* vol. 10, no. 5, pp. 715-727, 2008.

[96] A. Mogoutov and B. Kahane, "Data search strategy for science and technology emergence: A scalable and evolutionary query for nanotechnology tracking,"

Research Policy, vol. 36, no. 6, pp. 893-903, 2007.

[97] M. Zitt and E. Bassecoulard, "Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences," *Information Processing & Management,* vol. 42, no. 6, p. 1513–1531, 2006.

[98] K. Fitzgibbons and C. McNiven, "Towards a nanotechnology statistical framework," in *Blue Sky Indicators Conference II*, Ottawa, Canada, 2006.

[99] M. Alencar, A. Porter and A. Antunes, "Nanopatenting patterns in relation to product life cycle," *Technological Forecasting and Social Change,* vol. 74, no. 9, pp. 1661-1680, 2007.

[100] L. G. Zucker and M. R. Darby, "Socio-economic Impact of Nanoscale Science: Initial Results and NanoBank," National Bureau of Economic Research, 2005.

[101] ISI, "ISI web site on nanotechnology," The Research Services Group of Thomson Scientific, October 2002. [Online]. Available: http://www.esi-topics.com/nano/. [Accessed 7 May 2012].

[102] T. Braun, A. Schubert and S. Zsindely, "Nanoscience and nanotecnology on the balance," *Scientometrics,* vol. 38, no. 2, pp. 321-325, 1997-02-01.

[103] D. Marinova and M. McAleer, "Nanotechnology strength indicators: international rankings based on US patents," *Nanotechnology,* vol. 14, pp. R1-R7, 2003.

[104] R. N. Kostoff, J. A. Stump, D. Johnson, J. S. Murday, C. G. Lau and W. M. Tolles, "The structure and infrastructure of the global nanotechnology literature," *Journal of Nanoparticle Research,* vol. 8, no. 1, 2006.

[105] CABI, "CABI full text," CABI, Auguest 2011. [Online]. Available: http://www.cabi.org/default.aspx?site=170&page=1016&pid=2227. [Accessed 7 5

2012].

[106] M. E. Falagas, E. I. Pitsouni, G. A. Malietzis and G. Pappas, "Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses," *The FASEB Journal,* vol. 22, pp. 338-342, 2008.

[107] P. Jacso, "As we may search – Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases," *Current Science,* vol. 89, no. 9, pp. 1537-1547, 2005.

[108] L. Ford and L. H. O'Hara, "It's All Academic: Google Scholar, Scirus, and Windows Live Academic Search," *Journal of Library Administration,* vol. 46, no. 3-4, pp. 43-52, 2008.

[109] K. Faust, "Using Correspondence Analysis for Joint Displays of Affiliation Networks," in *Models and Methods in Social Network Analysis*, Cambridge, Cambridge University Press, 2005, pp. 117-147.

[110] H. Huisman and M. A. J. van Duijn, "Software for Social Network Analysis," in *Models and Methods in Social Network Analysis*, Cambridge , Cambridge University Press, 2005, pp. 270-316.

[111] M. S. Dahl and C. R. Pedersen, "Knowledge flows through informal contacts in industrial clusters: myth or reality?," *Research Policy,* vol. 33, no. 10, p. 1673–1686, 2004.

[112] A. Schiffauerova and C. Beaudry, "Gatekeepers of Canadian biotechnology clusters as interaction points between geographical and technological spaces of collaboration," *Journal of Engineering and Technology Management,* vol. 29, pp. 281-306, 2012.

[113] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publications of the*

Mathematical Institute of the Hungarian Academy of Sciences, vol. 5, p. 17–61, 1960.

[114] T. F. Bresnahan and M. Trajtenberg, "General Purpose Technologies: Engine of Growth?," *Journal of Econometrics,* vol. 65, pp. 83-108, 1995.

[115] J. Katz and B. Martin, "What is research collaboration?," *Research policy,* vol. 26, no. 1, pp. 1-18, 1997.

APPENDICES

Appendix I: List of Nanotechnology Keywords

Search Term	Search Queries
Nano* terms	"nano assembly", "nano computer", "nano cubic technology", "nano molecular machine", "nano optic", "nano optical tweezers", "nano warfare", “nanoarray”, "nanoassembler", "nanobarcode", "nanobarcodes particle", "nanobioprocess", "nanobot", "nanobotics", "nanobots", "nanobubble", "nanobusiness alliance", "nanobusiness company", "nanocatalysis", "nanoceramic", "nanochemistry", "nanochip", "nanocircle", "nanocluster", "nanocomputer", "nanocone", "nanocontact", "nanocrystal", "nanocrystal antenna", "nanodefense", "nanodentistry", "nanodetect", "nanodevice", "nanodiamond", "nanodisaster", "nanodot", "nanoelectrospray", "nanoengineering", "nanofacture", "nanofacty", "nanofiber", "nanofibre", "nanofiltration", "nanofluidic", "nanofoam", "nanogate", "nanogear", "nanogenomic", "nanoimaging", "nanoimprint lithography", "nanoimprint machine", "nanoimprinting", "nanolabel", "nanolithography", "nanomachine", "nanomagnet", "nanomanipulat", "nanomanipulation", "nanomanufacturing", "nanomaterial", "nanomechanical", "nanomot", "nanoparticles”,nanowire", "nanope", "nanope", "nanopharmaceutical", "nanophotonic", "nanophysic", "nanoplumbing", "nanoprism", "nano-ring", "nanoscale self assembly", "nanoscale synthesis", "nanoscience", "nanoscopic scale", "nanoscopic scale", "nanosens", "nanosheet", "nanoshell", "nanosource", "nanostructure", "nanostructured", "nanosurgery", "nanosystem", "nanotechism", "nanotechnology", "nanotube", "nanotube bundle", "nanowalker", "nanowetting"
Quantum terms	"quantum cascade laser", "quantum coherence", "quantum computation", "quantum compute", "quantum computer", "quantum

	computing", "quantum conduct", "quantum conductance", "quantum conductivity", "quantum confine", "quantum device", "quantum dot", "quantum gate", "quantum information", "quantum information process", "quantum mirage", "quantum nanophysics", "quantum nanomechanics", "quantum system", "quantum well"
Molecular* terms	"molecular assembler", "molecular machine", "molecular nanogenerat", "molecular nanotechnology", "molecular robotic", "molecular scale manufacturing", "molecular systems engineering", "molecular technology"
Self assembly terms	"fluidic self assembly", "nanoscale self assembly", "self assembled"
Atomic terms	"atomic manipulation", "atomic nanostructure"
Other terms	"biofabrication", "biomedical nanotechnology", "biomimetic synthesis", "biomolecular assembly", "biomolecular nanoscale computing", "biomolecular nanotechnology", "bionems", "brownian assembly", "buckminsterfullerene", "buckyball", "buckytube", "c60 molecule", "carbon nanotubes", "conductance quantization", "dna chip", "electron beam lithography", "epitaxial film", "epitaxy", "fat fingers problem", "ganic led", "glyconanotechnology", "grey.goo", "immune machine", "khaki goo", "laser tweezer", "limited assembler", "military nanotech.", "moletronic", "naneplicat", "nanite", "optical trapping", "protein design", "protein engineering", "proximal probe", "rotaxane", "single cell manipulation", "spin coating", "stewart platfm", "sticky fingers problem", "textronic", "universal assembler", "utility fog", "zettatechnology"

Appendix II: Internal Collaboration Patterns of Regions in the Network of Articles

Quebec	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
Size of the Sub-Network													
# of Authors	349	452	541	598	644	718	838	953	1067	1146	1322	1467	1767
Percentage in the Network	0.39%	0.45%	0.49%	0.49%	0.50%	0.51%	0.53%	0.53%	0.53%	0.50%	0.49%	0.47%	0.50%
Collaborations inside Quebec													
Total # of Pairs	204	327	417	523	557	684	785	861	1083	1143	1445	1567	2409
Total # of Collaborations	502	812	1098	1339	1416	1648	1892	2114	2641	2731	3397	3716	5951
Avg. Collaboration / Author	1.4384	1.7965	2.0296	2.2408	2.1988	2.2953	2.2578	2.2183	2.4761	2.3839	2.5703	2.5331	3.3684
Max Collaboration / Author	24	29	40	38	33	32	47	44	49	38	44	32	47
Collaborators inside Quebec													
Avg. Collaborators / Author	1.1691	1.4469	1.5416	1.7492	1.7298	1.9053	1.8735	1.8069	2.0300	1.9948	2.1861	2.1363	2.7267
Max Collaborator / Author	12	17	16	16	12	12	16	16	19	18	20	19	21
Repeated Collaborations													
Repeated Collaborations	35	59	78	91	102	96	107	116	141	150	160	188	392
Avg. of Repeated Collaboration	0.5931	0.6792	0.7375	0.7742	0.7314	0.734	0.7446	0.7692	0.8154	0.801	0.8321	0.8759	0.9932
Max Repeated Collaboration	6	6	9	8	9	8	10	12	12	12	10	7	7
Isolated Authors													
Isolated Authors	203	229	270	288	320	333	372	423	443	489	517	525	521
% of Isolated Authors	58.17%	50.66%	49.91%	48.16%	49.69%	46.38%	44.39%	44.39%	41.52%	42.67%	39.11%	35.79%	29.49%
Connected Component Results													
Largest Component (LC) Size	16	23	24	24	31	35	29	29	49	66	191	202	343
Share of LC	4.58%	5.09%	4.44%	4.01%	4.81%	4.87%	3.46%	3.04%	4.59%	5.76%	14.45%	13.77%	19.41%
Diameter in LC	4	4	6	8	6	7	6	9	8	7	16	17	18
Size of Second LC	11	13	18	19	20	24	24	22	25	28	44	23	53
Share of Second LC	3.15%	2.88%	3.33%	3.18%	3.11%	3.34%	2.86%	2.31%	2.34%	2.44%	3.33%	1.57%	3.00%
Avg. Component Size	1.4664	1.5972	1.6149	1.6893	1.6727	1.7385	1.7495	1.7486	1.8621	1.8664	2.0401	2.1108	2.4576

Rest of Canada	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
Size of the Sub-Network													
# of Authors	1,296	1,723	2,113	2,249	2,405	2,617	2,970	3,343	3,743	4,258	4,929	5,627	6383
Percentage in the Network	1.45%	1.72%	1.93%	1.85%	1.87%	1.85%	1.87%	1.87%	1.87%	1.84%	1.83%	1.82%	1.79%
Collaboration Pairs													
Total # of Pairs	1,091	1,535	2,107	2,248	2,358	2,442	2,981	3,463	4,121	5,211	6,762	8,945	10,764
Total # of Collaborations	3,273	4,209	5,634	5,981	6,234	6,177	7,414	8,564	10,298	13,256	17,603	22,829	27,752
Avg. Collaboration / Author	2.5262	2.4434	2.6664	2.6598	2.5921	2.3607	2.4963	2.5618	2.7513	3.1132	3.5715	4.0572	4.3478
Max Collaboration / Author	93	96	126	112	109	109	109	82	104	101	94	116	165
Collaborators inside Quebec													
Avg. Collaborators / Author	1.6836	1.7818	1.9943	1.9991	1.9609	1.8663	2.0074	2.0718	2.2020	2.4476	2.7438	3.1793	3.3727
Max Collaborator / Author	32	34	51	46	40	36	42	38	43	39	40	53	56
Repeated Collaborations													
Repeated Collaborations	249	293	390	392	396	334	373	435	584	729	1,038	1,322	1,628
Avg. of Repeated Collaboration	0.8056	0.7911	0.8268	0.8221	0.7979	0.747	0.7764	0.805	0.8605	0.9497	1.0422	1.0947	1.146
Max Repeated Collaboration	16	17	17	17	15	16	17	13	11	13	20	19	19
Isolated Authors													
Isolated Authors	685	835	931	992	1,097	1,252	1,341	1,459	1,528	1,618	1,656	1,705	1,703
% of Isolated Authors	52.85%	48.46%	44.06%	44.11%	45.61%	47.84%	45.15%	43.64%	40.82%	38.00%	33.60%	30.30%	26.68%
Connected Component Results													
Largest Component (LC) Size	179	209	236	251	204	184	215	242	347	479	873	1,364	1,492
Share of LC	13.81%	12.13%	11.17%	11.16%	8.48%	7.03%	7.24%	7.24%	9.27%	11.25%	17.71%	24.24%	23.37%
Diameter in LC	10	13	10	10	11	11	12	12	16	17	20	34	23
Size of Second LC	84	115	65	70	64	82	120	149	157	128	84	37	217
Share of Second LC	6.48%	6.67%	3.08%	3.11%	2.66%	3.13%	4.04%	4.46%	4.19%	3.01%	1.70%	0.66%	3.40%
Avg. Component Size	1.6119	1.6843	1.7741	1.7807	1.7453	1.7093	1.7978	1.8278	1.9254	2.0276	2.2303	2.4192	2.6674

Canada	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
Size of the Sub-Network													
# of Authors	1,645	2,175	2,654	2,847	3,049	3,335	3,808	4,296	4,810	5,404	6,251	7,094	8,150
Percentage in the Network	1.84%	2.17%	2.42%	2.34%	2.37%	2.36%	2.40%	2.40%	2.40%	2.34%	2.32%	2.29%	2.29%
Collaboration Pairs													
Total # of Pairs	1,510	2,163	2,874	3,194	3,342	3,584	4,310	4,894	5,829	6,971	8,997	11,424	14,401
Total # of Collaborations	4,330	5,770	7,650	8,380	8,712	8,896	10,558	12,010	14,472	17,548	23,012	28,928	36,802
Avg. Collaboration / Author	2.6322	2.6529	2.8824	2.9434	2.8573	2.6675	2.7726	2.7956	3.0087	3.2472	3.6813	4.0778	4.5156
Max Collaboration / Author	99	101	135	122	126	120	117	82	107	102	103	130	172
Collaborators inside Quebec													
Avg. Collaborators / Author	1.8359	1.9890	2.1658	2.2438	2.1922	2.1493	2.2637	2.2784	2.4237	2.5799	2.8786	3.2207	3.5340
Max Collaborator / Author	35	39	56	52	49	45	50	42	45	47	48	54	56
Repeated Collaborations													
Repeated Collaborations	326	401	536	556	578	487	543	608	806	961	1,300	1,653	2,190
Avg. of Repeated Collaboration	0.8225	0.8294	0.8685	0.8672	0.8321	0.785	0.8117	0.8424	0.899	0.9634	1.0426	1.0923	1.1523
Max Repeated Collaboration	16	17	17	17	15	16	17	13	12	14	20	19	19
Isolated Authors													
Isolated Authors	829	989	1,116	1,190	1,328	1,493	1,611	1,759	1,836	1,968	2,014	2,073	2,072
% of Isolated Authors	50.40%	45.47%	42.05%	41.80%	43.56%	44.77%	42.31%	40.95%	38.17%	36.42%	32.22%	29.22%	25.42%
Connected Component Results													
Largest Component (LC) Size	213	264	292	336	300	319	275	395	541	697	1,311	1,776	2,407
Share of LC	12.95%	12.14%	11.00%	11.80%	9.84%	9.57%	7.22%	9.19%	11.25%	12.90%	20.97%	25.04%	29.53%
Diameter in LC	10	13	11	14	12	18	19	19	21	21	21	25	30
Size of Second LC	103	135	91	80	75	101	268	232	189	139	85	72	244
Share of Second LC	6.26%	6.21%	3.43%	2.81%	2.46%	3.03%	7.04%	5.40%	3.93%	2.57%	1.36%	1.01%	2.99%
Avg. Component Size	1.6769	1.7755	1.8572	1.8867	1.8445	1.8184	1.8993	1.9282	2.0381	2.1184	2.3368	2.5183	2.8171

The United States	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
Size of the Sub-Network													
# of Authors	19,859	25,563	31,091	33,602	34,811	37,728	42,386	47,486	52,349	59,424	67,638	75,857	84,989
Percentage in the Network	22.16%	25.54%	28.35%	27.64%	27.05%	26.65%	26.73%	26.54%	26.10%	25.69%	25.08%	24.50%	23.89%
Collaboration Pairs													
Total # of Pairs	33,836	63,494	77,433	81,665	64,547	67,185	76,284	93,020	110,168	134,530	163,599	192,960	229,402
Total # of Collaborations	97,074	161,555	192,524	203,500	164,962	168,481	187,769	231,774	278,941	348,783	423,954	494,428	572,290
Avg. Collaboration / Author	4.8882	6.3199	6.1923	6.0562	4.7388	4.4657	4.4300	4.8809	5.3285	5.8694	6.2680	6.5179	6.7337
Max Collaboration / Author	268	293	264	323	247	264	253	244	230	288	366	284	314
Collaborators inside Quebec													
Avg. Collaborators / Author	3.4076	4.9676	4.9811	4.8607	3.7084	3.5615	3.5995	3.9178	4.2090	4.5278	4.8375	5.0875	5.3984
Max Collaborator / Author	78	174	172	172	129	142	129	126	95	114	138	127	186
Repeated Collaborations													
Repeated Collaborations	7,439	9,276	9,996	10,579	9,676	9,348	9,937	12,703	16,069	21,352	25,714	29,649	31,907
Avg. of Repeated Collaboration	1.2004	1.1814	1.1495	1.1483	1.0828	1.0358	1.0243	1.0961	1.1715	1.2621	1.3098	1.3313	1.3306
Max Repeated Collaboration	54	46	40	44	36	46	50	52	66	81	76	80	62
Isolated Authors													
Isolated Authors	6,932	7,616	8,533	9,362	10,359	11,466	12,621	12,855	13,277	13,990	14,380	14,451	14,027
% of Isolated Authors	34.91%	29.79%	27.45%	27.86%	29.76%	30.39%	29.78%	27.07%	25.36%	23.54%	21.26%	19.05%	16.50%
Connected Component Results													
Largest Component (LC) Size	7,071	10,032	12,291	13,681	12,916	13,272	14,785	18,560	22,712	27,531	33,760	40,688	47,459
Share of LC	35.61%	39.24%	39.53%	40.71%	37.10%	35.18%	34.88%	39.09%	43.39%	46.33%	49.91%	53.64%	55.84%
Diameter in LC	27	27	24	31	23	27	28	24	28	25	26	25	24
Size of Second LC	117	175	173	173	61	44	80	73	58	68	111	73	68
Share of Second LC	0.59%	0.68%	0.56%	0.51%	0.18%	0.12%	0.19%	0.15%	0.11%	0.11%	0.16%	0.10%	0.08%
Avg. Component Size	2.3322	2.6414	2.8119	2.8048	2.6247	2.5613	2.5986	2.8222	3.0209	3.2447	3.545	3.9007	4.3201

Appendix III: Collaboration Patterns between Regions in the Network of Articles

QC – Rest of CA	**94-96**	**95-97**	**96-98**	**97-99**	**98-00**	**99-01**	**00-02**	**01-03**	**02-04**	**03-05**	**04-06**	**05-07**	**06-08**
Authors of Quebec													
# of Authors	349	452	541	598	644	718	838	953	1,067	1,146	1,322	1,467	1,767
# of Isolated	260	321	377	407	461	525	609	691	754	852	985	1,084	1,277
# of Gatekeepers	89	131	164	191	183	193	229	262	313	294	337	383	490
% of Gatekeepers	25.50%	28.98%	30.31%	31.94%	28.42%	26.88%	27.33%	27.49%	29.33%	25.65%	25.49%	26.11%	27.73%
Authors of the Rest of Canada													
# of Authors	1,296	1,723	2,113	2,249	2,405	2,617	2,970	3,343	3,743	4,258	4,929	5,627	6,383
# of Isolated	1,173	1,550	1,921	2,038	2,177	2,372	2,676	3,055	3,442	3,942	4,523	5,180	5,858
# of Gatekeepers	123	173	192	211	228	245	294	288	301	316	406	447	525
% of Gatekeepers	9.49%	10.04%	9.09%	9.38%	9.48%	9.36%	9.90%	8.62%	8.04%	7.42%	8.24%	7.94%	8.22%
Collaborations													
# of Pairs	215	301	350	423	427	458	544	570	625	617	790	912	1,228
# of Collaborations	554	748	918	1,058	1,062	1,070	1,252	1,332	1,532	1,560	2,010	2,382	3,098
Avg. of Collaboration(QC)	1.5874	1.6549	1.6969	1.7692	1.6491	1.4903	1.4940	1.3977	1.4358	1.3613	1.5204	1.6237	1.7533
Avg. of Collaboration (Rest of CA)	0.4275	0.4341	0.4345	0.4704	0.4416	0.4089	0.4215	0.3984	0.4093	0.3664	0.4078	0.4233	0.4854
Max of Collaboration / Author (QC)	16	20	20	20	22	22	24	22	33	29	41	48	56
Max of Collaboration / Author (Rest of CA)	13	10	14	14	19	21	15	15	17	31	26	26	26
Collaborators													
Avg. of Collaborators / Author (QC)	0.616	0.6659	0.647	0.7074	0.663	0.6379	0.6492	0.5981	0.5858	0.5384	0.5976	0.6217	0.695
Max of Collaborator / Author (QC)	11	14	11	14	17	17	17	15	14	11	23	24	25

QC – Rest of CA	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
Avg. of Collaborators / Author (Rest of CA)	0.1659	0.1747	0.1656	0.1881	0.1775	0.175	0.1832	0.1705	0.167	0.1449	0.1603	0.1621	0.1924
Max of Collaborator / Author (Rest of CA)	8	8	9	10	10	10	9	9	11	10	11	11	13
Repeated Collaborations													
# of Repeated Collaborations	42	49	68	73	80	57	63	57	81	82	102	143	170
Avg. of Repeated Collaboration / Author	0.183	0.1903	0.1933	0.1932	0.1804	0.1625	0.1686	0.1615	0.1709	0.1564	0.1656	0.1672	0.1737
Max of Repeated Collaboration / Author	5	8	7	7	5	5	4	7	8	14	14	17	10

QC - US	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
Authors of Quebec													
# of Authors	349	452	541	598	644	718	838	953	1,067	1,146	1,322	1,467	1,767
# of Isolated	244	305	371	404	446	493	586	672	754	854	973	1,044	1,236
# of Gatekeepers	105	147	170	194	198	225	252	281	313	292	349	423	531
% of Gatekeepers	30.09%	32.52%	31.42%	32.44%	30.75%	31.34%	30.07%	29.49%	29.33%	25.48%	26.40%	28.83%	30.05%
Authors of the United States													
# of Authors	19,859	25,563	31,091	33,602	34,811	37,728	42,386	47,486	52,349	59,424	67,638	75,857	84,989
# of Isolated	19,689	25,146	30,614	33,073	34,450	37,309	41,916	46,931	51,665	58,707	66,736	74,869	83,913
# of Gatekeepers	170	417	477	529	361	419	470	555	684	717	902	988	1,076
% of Gatekeepers	0.86%	1.63%	1.53%	1.57%	1.04%	1.11%	1.11%	1.17%	1.31%	1.21%	1.33%	1.30%	1.27%
Collaborations													

QC - US

	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
# of Pairs	222	681	756	839	495	619	703	820	954	942	1,170	1,356	1,830
# of Collaborations	550	1,514	1,704	1,948	1,222	1,474	1,702	1,990	2,304	2,316	2,870	3,316	4,314
Avg. of Collaboration(QC)	1.5759	3.3496	3.1497	3.2575	1.8975	2.0529	2.0310	2.0881	2.1593	2.0209	2.1710	2.2604	2.4414
Avg. of Collaboration (US)	0.0277	0.0592	0.0548	0.0580	0.0351	0.0391	0.0402	0.0419	0.0440	0.0390	0.0424	0.0437	0.0508
Max of Collaboration / Author (QC)	11	173	173	173	34	34	40	61	56	37	57	71	67
Max of Collaboration / Author (US)	23	13	27	30	17	15	17	21	15	10	14	21	24
Collaborators													
Avg. of Collaborators / Author (QC)	0.6361	1.5066	1.3974	1.403	0.7686	0.8621	0.8389	0.8604	0.8941	0.822	0.885	0.9243	1.0357
Max of Collaborator / Author (QC)	9	173	173	173	19	18	21	23	23	22	34	38	67
Avg. of Collaborators / Author (US)	0.0112	0.0266	0.0243	0.025	0.0142	0.0164	0.0166	0.0173	0.0182	0.0159	0.0173	0.0179	0.0215
Max of Collaborator / Author (US)	7	7	10	10	9	9	9	8	8	8	10	14	18
Repeated Collaborations													
# of Repeated Collaborations	31	49	61	81	74	74	85	97	117	145	175	172	175
Avg. of Repeated Collaboration / Author	0.0178	0.0259	0.0242	0.0262	0.0204	0.0211	0.0215	0.0223	0.024	0.0219	0.0237	0.0239	0.024
Max of Repeated Collaboration / Author	8	6	7	8	5	10	13	16	11	7	9	11	13

QC –
Rest of the Word

	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
Authors of Quebec													

QC – Rest of the Word	**94-96**	**95-97**	**96-98**	**97-99**	**98-00**	**99-01**	**00-02**	**01-03**	**02-04**	**03-05**	**04-06**	**05-07**	**06-08**
# of Authors	349	452	541	598	644	718	838	953	1,067	1,146	1,322	1,467	1,767
# of Isolated	155	226	270	286	319	362	434	505	598	625	704	718	855
# of Gatekeepers	194	226	271	312	325	356	404	448	469	521	618	749	912
% of Gatekeepers	55.59%	50.00%	50.09%	52.17%	50.47%	49.58%	48.21%	47.01%	43.96%	45.46%	46.75%	51.06%	51.61%
Authors of the Rest of the World													
# of Authors	68,122	72,351	75,909	85,115	90,808	100,488	112,379	127,108	143,436	166,487	195,794	226,619	262,628
# of Isolated	67,595	71,753	75,256	84,371	90,040	99,648	111,413	125,981	142,179	165,033	194,001	224,449	260,073
# of Gatekeepers	527	598	653	744	768	840	966	1,127	1,257	1,454	1,793	2,170	2,555
% of Gatekeepers	0.77%	0.83%	0.86%	0.87%	0.85%	0.84%	0.86%	0.89%	0.88%	0.87%	0.92%	0.96%	0.97%
Collaborations													
# of Pairs	694	800	923	1,075	1,078	1,207	1,377	1,596	1,665	1,879	2,335	2,859	4,341
# of Collaborations	1,614	1,860	2,156	2,532	2,542	2,904	3,366	4,016	4,356	4,942	5,970	7,198	10,398
Avg. of Collaboration(QC)	4.6246	4.1150	3.9852	4.2341	3.9472	4.0446	4.0167	4.2141	4.0825	4.3124	4.5159	4.9066	5.8846
Avg. of Collaboration (World)	0.0237	0.0257	0.0284	0.0297	0.0280	0.0289	0.0300	0.0316	0.0304	0.0297	0.0305	0.0318	0.0396
Max of Collaboration / Author (QC)	40	24	24	43	33	75	88	111	106	109	100	88	214
Max of Collaboration / Author (World)	8	12	19	21	19	15	18	23	18	19	14	16	18
Collaborators													

QC – Rest of the Word

	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
Avg. of Collaborators / Author (QC)	1.9885	1.7699	1.7061	1.7977	1.6739	1.6811	1.6432	1.6747	1.5604	1.6396	1.7663	1.9489	2.4567
Max of Collaborator / Author (QC)	28	22	22	23	21	39	48	48	57	48	55	64	214
Avg. of Collaborators / Author (World)	0.0102	0.0111	0.0122	0.0126	0.0119	0.012	0.0123	0.0126	0.0116	0.0113	0.0119	0.0126	0.0165
Max of Collaborator / Author (World)	6	6	8	10	10	10	7	8	8	8	11	11	18
Repeated Collaborations													
# of Repeated Collaborations	90	103	113	132	127	147	170	205	275	335	392	429	497
Avg. of Repeated Collaboration / Author	0.0129	0.0138	0.0148	0.0152	0.0148	0.0149	0.0155	0.0164	0.0164	0.0162	0.0165	0.017	0.0172
Max of Repeated Collaboration / Author	7	6	10	13	13	13	18	23	18	11	10	13	11

The Rest of Canada – US

	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
Authors of the Rest of Canada													
# of Authors	1,296	1,723	2,113	2,249	2,405	2,617	2,970	3,343	3,743	4,258	4,929	5,627	6,383
# of Isolated	839	1,043	1,246	1,360	1,519	1,732	1,999	2,217	2,432	2,681	3,100	3,607	4,141
# of Gatekeepers	457	680	867	889	886	885	971	1,126	1,311	1,577	1,829	2,020	2,242
% of Gatekeepers	35.26%	39.47%	41.03%	39.53%	36.84%	33.82%	32.69%	33.68%	35.03%	37.04%	37.11%	35.90%	35.12%
Authors of the United States													
# of Authors	19,859	25,563	31,091	33,602	34,811	37,728	42,386	47,486	52,349	59,424	67,638	75,857	84,989
# of Isolated	19,066	24,205	29,417	31,800	33,188	36,050	40,608	45,439	50,041	56,667	64,514	72,425	81,243

The Rest of Canada – US	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
# of Gatekeepers	793	1,358	1,674	1,802	1,623	1,678	1,778	2,047	2,308	2,757	3,124	3,432	3,746
% of Gatekeepers	3.99%	5.31%	5.38%	5.36%	4.66%	4.45%	4.19%	4.31%	4.41%	4.64%	4.62%	4.52%	4.41%
Collaborations													
# of Pairs	1,273	3,147	3,708	3,803	2,515	2,540	2,739	3,216	3,695	4,593	5,397	6,042	6,899
# of Collaborations	3,380	7,366	8,728	9,036	6,510	6,372	6,738	7,830	9,240	11,494	13,310	14,882	16,686
Avg. of Collaboration(REST)	2.6080	4.2751	4.1306	4.0178	2.7069	2.4348	2.2687	2.3422	2.4686	2.6994	2.7003	2.6447	2.6141
Avg. of Collaboration (US)	0.1702	0.2882	0.2807	0.2689	0.1870	0.1689	0.1590	0.1649	0.1765	0.1934	0.1968	0.1962	0.1963
Max of Collaboration / Author (REST)	49	173	173	173	89	73	92	79	58	67	53	69	91
Max of Collaboration / Author (US)	40	36	43	46	36	29	19	34	45	40	51	56	73
Collaborators													
Avg. of Collaborators / Author (REST)	0.9823	1.8265	1.7549	1.691	1.0457	0.9706	0.9222	0.962	0.9872	1.0787	1.0949	1.0738	1.0808
Max of Collaborator / Author (REST)	26	173	173	173	43	36	36	38	33	29	40	46	67
Avg. of Collaborators / Author (US)	0.0641	0.1231	0.1193	0.1132	0.0722	0.0673	0.0646	0.0677	0.0706	0.0773	0.0798	0.0796	0.0812
Max of Collaborator / Author (US)	18	23	30	25	20	15	14	19	21	22	33	32	33
Repeated Collaborations													
# of Repeated Collaborations	240	331	409	412	425	357	351	404	541	647	763	833	848
Avg. of Repeated Collaboration / Author	0.0854	0.0994	0.1005	0.0999	0.0922	0.0846	0.0791	0.0815	0.0871	0.0924	0.0913	0.0893	0.0866
Max of Repeated Collaboration / Author	10	8	11	12	10	12	13	9	10	14	12	12	12

Rest of Canada – Rest of the World	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
Authors of the Rest of Canada													
# of Authors	1,296	1,723	2,113	2,249	2,405	2,617	2,970	3,343	3,743	4,258	4,929	5,627	6,383
# of Isolated	584	853	1,095	1,173	1,280	1,423	1,615	1,834	2,004	2,243	2,486	2,875	3,190
# of Gatekeepers	712	870	1,018	1,076	1,125	1,194	1,355	1,509	1,739	2,015	2,443	2,752	3,193
% of Gatekeepers	54.94%	50.49%	48.18%	47.84%	46.78%	45.62%	45.62%	45.14%	46.46%	47.32%	49.56%	48.91%	50.02%
Authors of the United States													
# of Authors	68,122	72,351	75,909	85,115	90,808	100,488	112,379	127,108	143,436	166,487	195,794	226,619	262,628
# of Isolated	66,306	70,390	73,668	82,576	88,006	97,508	109,113	123,456	139,142	161,544	189,961	220,138	254,758
# of Gatekeepers	1,816	1,961	2,241	2,539	2,802	2,980	3,266	3,652	4,294	4,943	5,833	6,481	7,870
% of Gatekeepers	2.67%	2.71%	2.95%	2.98%	3.09%	2.97%	2.91%	2.87%	2.99%	2.97%	2.98%	2.86%	3.00%
Collaborations													
# of Pairs	2,792	3,163	3,512	3,891	4,053	4,347	4,729	5,376	6,330	7,504	9,104	10,400	13,238
# of Collaborations	7,208	7,780	8,802	9,800	10,204	10,672	11,532	13,098	15,862	19,372	23,492	27,024	33,480
Avg. of Collaboration(REST)	5.5617	4.5154	4.1656	4.3575	4.2428	4.0780	3.8828	3.9180	4.2378	4.5496	4.7661	4.8026	5.2452
Avg. of Collaboration (US)	0.1058	0.1075	0.1160	0.1151	0.1124	0.1062	0.1026	0.1030	0.1106	0.1164	0.1200	0.1192	0.1275
Max of Collaboration / Author (REST)	135	75	67	93	93	93	88	81	157	187	127	117	214
Max of Collaboration / Author (US)	79	91	88	87	81	48	38	31	39	53	51	47	97

Rest of Canada – Rest of the World	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
Collaborators													
Avg. of Collaborators / Author (REST)	2.1543	1.8358	1.6621	1.7301	1.6852	1.6611	1.5923	1.6081	1.6912	1.7623	1.847	1.8482	2.0739
Max of Collaborator / Author (REST)	63	63	63	93	93	93	48	48	50	59	68	72	214
Avg. of Collaborators / Author (US)	0.041	0.0437	0.0463	0.0457	0.0446	0.0433	0.0421	0.0423	0.0441	0.0451	0.0465	0.0459	0.0504
Max of Collaborator / Author (US)	30	33	39	38	34	28	14	16	19	26	27	40	40
Repeated Collaborations													
# of Repeated Collaborations	446	434	513	567	619	576	581	697	929	1194	1470	1710	1946
Avg. of Repeated Collaboration / Author	0.0506	0.0501	0.0554	0.0553	0.0556	0.052	0.0421	0.0512	0.0548	0.0562	0.0572	0.0561	0.057
Max of Repeated Collaboration / Author	18	13	12	12	15	14	14	11	17	22	17	17	21

CA – US	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
Authors of Canada													
# of Authors	1,645	2,175	2,654	2,847	3,049	3,335	3,808	4,296	4,810	5,404	6,251	7,094	8,150
# of Isolated	1,083	1,348	1,617	1,764	1,965	2,225	2,585	2,889	3,186	3,535	4,073	4,651	5,377
# of Gatekeepers	562	827	1,037	1,083	1,084	1,110	1,223	1,407	1,624	1,869	2,178	2,443	2,773
% of Gatekeepers	34.16%	38.02%	39.07%	38.04%	35.55%	33.28%	32.12%	32.75%	33.76%	34.59%	34.84%	34.44%	34.02%
Authors of the United States													
# of Authors	19,859	25,563	31,091	33,602	34,811	37,728	42,386	47,486	52,349	59,424	67,638	75,857	84,989

CA – US

	94-96	95-97	96-98	97-99	98-00	99-01	00-02	01-03	02-04	03-05	04-06	05-07	06-08
# of Isolated	18,938	24,038	29,205	31,557	32,928	35,736	40,242	45,003	49,526	56,135	63,845	71,660	80,452
# of Gatekeepers	921	1,525	1,886	2,045	1,883	1,992	2,144	2,483	2,823	3,289	3,793	4,197	4,537
% of Gatekeepers	4.64%	5.97%	6.07%	6.09%	5.41%	5.28%	5.06%	5.23%	5.39%	5.53%	5.61%	5.53%	5.34%
Collaborations													
# of Pairs	1,495	3,828	4,464	4,642	3,010	3,159	3,442	4,036	4,649	5,535	6,567	7,398	8,729
# of Collaborations	3,930	8,880	10,432	10,984	7,732	7,846	8,440	9,820	11,544	13,810	16,180	18,198	21,000
Avg. of Collaboration(CA)	2.3891	4.0828	3.9307	3.8581	2.5359	2.3526	2.2164	2.2858	2.4000	2.5555	2.5884	2.5653	2.5767
Avg. of Collaboration (US)	0.1979	0.3474	0.3355	0.3269	0.2221	0.2080	0.1991	0.2068	0.2205	0.2324	0.2392	0.2399	0.2471
Max of Collaboration / Author (CA)	49	173	173	173	89	73	92	79	58	67	57	71	91
Max of Collaboration / Author (US)	40	36	46	48	39	32	24	36	46	48	59	67	84
Collaborators													
Avg. of Collaborators / Author (CA)	0.9088	1.76	1.682	1.6305	0.9872	0.9472	0.9039	0.9395	0.9665	1.0242	1.0506	1.0429	1.071
Max of Collaborator / Author (CA)	26	173	173	173	43	36	36	38	33	29	40	46	67
Avg. of Collaborators / Author (US)	0.0753	0.1497	0.1436	0.1381	0.0865	0.0837	0.0812	0.085	0.0888	0.0931	0.0971	0.0975	0.1027
Max of Collaborator / Author (US)	18	27	34	26	21	16	19	19	26	23	34	33	33
Repeated Collaborations													
# of Repeated Collaborations	271	380	470	493	499	431	436	501	658	792	938	1005	1023
Avg. of Repeated Collaboration / Author	0.0983	0.1125	0.1134	0.1147	0.1068	0.1002	0.0953	0.0984	0.1046	0.1079	0.1083	0.1066	0.1036
Max of Repeated Collaboration / Author	10	8	11	12	10	12	13	16	11	14	12	12	13

CA – Rest of the World	**94-96**	**95-97**	**96-98**	**97-99**	**98-00**	**99-01**	**00-02**	**01-03**	**02-04**	**03-05**	**04-06**	**05-07**	**06-08**
Authors of Canada													
# of Authors	1,645	2,175	2,654	2,847	3,049	3,335	3,808	4,296	4,810	5,404	6,251	7,094	8,150
# of Isolated	739	1,079	1,365	1,459	1,599	1,785	2,049	2,339	2,602	2,868	3,190	3,593	4,045
# of Gatekeepers	906	1,096	1,289	1,388	1,450	1,550	1,759	1,957	2,208	2,536	3,061	3,501	4,105
% of Gatekeepers	55.08%	50.39%	48.57%	48.75%	47.56%	46.48%	46.19%	45.55%	45.90%	46.93%	48.97%	49.35%	50.37%
Authors of the Rest of the World													
# of Authors	68,122	72,351	75,909	85,115	90,808	100,488	112,379	127,108	143,436	166,487	195,794	226,619	262,628
# of Isolated	65,917	69,953	73,176	82,013	87,402	96,847	108,373	122,582	138,141	160,319	188,452	218,303	252,832
# of Gatekeepers	2,205	2,398	2,733	3,102	3,406	3,641	4,006	4,526	5,295	6,168	7,342	8,316	9,796
% of Gatekeepers	3.24%	3.31%	3.60%	3.64%	3.75%	3.62%	3.56%	3.56%	3.69%	3.70%	3.75%	3.67%	3.73%
Collaborations													
# of Pairs	3,486	3,963	4,435	4,966	5,131	5,554	6,106	6,972	7,995	9,383	11,439	13,259	17,579
# of Collaborations	8,822	9,640	10,958	12,332	12,746	13,576	14,898	17,114	20,218	24,314	29,462	34,222	43,878
Avg. of Collaboration(CA)	5.3629	4.4322	4.1289	4.3316	4.1804	4.0708	3.9123	3.9837	4.2033	4.4993	4.7132	4.8241	5.3838
Avg. of Collaboration (World)	0.1295	0.1332	0.1444	0.1449	0.1404	0.1351	0.1326	0.1346	0.1410	0.1460	0.1505	0.1510	0.1671
Max of Collaboration / Author (CA)	135	75	67	93	93	93	88	111	157	187	127	117	241
Max of Collaboration / Author (World)	79	92	94	95	89	50	38	37	44	53	56	47	99
Collaborators													

CA – Rest of the World	**94-96**	**95-97**	**96-98**	**97-99**	**98-00**	**99-01**	**00-02**	**01-03**	**02-04**	**03-05**	**04-06**	**05-07**	**06-08**
Avg. of Collaborators / Author (CA)	2.1191	1.8221	1.6711	1.7443	1.6828	1.6654	1.6035	1.6229	1.6622	1.7363	1.8299	1.869	2.1569
Max of Collaborator / Author (CA)	63	63	63	93	93	93	48	48	57	59	68	72	214
Avg. of Collaborators / Author (World)	0.0512	0.0548	0.0584	0.0583	0.0565	0.0553	0.0543	0.0549	0.0557	0.0564	0.0584	0.0585	0.0669
Max of Collaborator / Author (World)	30	33	42	43	39	30	18	16	20	27	27	40	40
Repeated Collaborations													
# of Repeated Collaborations	536	537	626	699	746	723	751	902	1204	1529	1862	2139	2443
Avg. of Repeated Collaboration / Author	0.0609	0.061	0.0671	0.0674	0.0677	0.0644	0.0638	0.0646	0.0684	0.0701	0.0713	0.0706	0.0708
Max of Repeated Collaboration / Author	18	13	12	13	15	14	18	23	18	22	17	17	21

US – Rest of the World	**94-96**	**95-97**	**96-98**	**97-99**	**98-00**	**99-01**	**00-02**	**01-03**	**02-04**	**03-05**	**04-06**	**05-07**	**06-08**
Authors of the United States													
# of Authors	19,859	25,563	31,091	33,602	34,811	37,728	42,386	47,486	52,349	59,424	67,638	75,857	84,989
# of Isolated	9,009	12,313	15,880	17,038	18,241	20,313	23,296	25,483	27,134	29,814	33,230	37,217	41,574
# of Gatekeepers	10,850	13,250	15,211	16,564	16,570	17,415	19,090	22,003	25,215	29,610	34,408	38,640	43,415
% of Gatekeepers	54.64%	51.83%	48.92%	49.29%	47.60%	46.16%	45.04%	46.34%	48.17%	49.83%	50.87%	50.94%	51.08%
Authors of the Rest of the World													
# of Authors	68,122	72,351	75,909	85,115	90,808	100,488	112,379	127,108	143,436	166,487	195,794	226,619	262,628

US – Rest of the World	**94-96**	**95-97**	**96-98**	**97-99**	**98-00**	**99-01**	**00-02**	**01-03**	**02-04**	**03-05**	**04-06**	**05-07**	**06-08**
# of Isolated	49,657	52,881	55,937	62,819	67,851	76,108	86,046	97,120	108,944	125,931	147,895	171,548	199,529
# of Gatekeepers	18,465	19,470	19,972	22,296	22,957	24,380	26,333	29,988	34,492	40,556	47,899	55,071	63,099
% of Gatekeepers	27.11%	26.91%	26.31%	26.20%	25.28%	24.26%	23.43%	23.59%	24.05%	24.36%	24.46%	24.30%	24.03%
Collaborations													
# of Pairs	44,381	51,405	52,965	58,845	55,806	58,809	63,948	75,667	88,736	107,041	128,154	145,950	182,603
# of Collaborations	114,738	127,162	131,920	147,146	141,778	148,452	161,392	192,466	228,184	280,690	337,254	381,528	460,006
Avg. of Collaboration(CA)	5.7776	4.9745	4.2430	4.3791	4.0728	3.9348	3.8077	4.0531	4.3589	4.7235	4.9862	5.0296	5.4125
Avg. of Collaboration (US)	1.6843	1.7576	1.7379	1.7288	1.5613	1.4773	1.4361	1.5142	1.5908	1.6860	1.7225	1.6836	1.7515
Max of Collaboration / Author (CA)	373	222	229	250	293	255	350	387	336	338	391	367	411
Max of Collaboration / Author (US)	128	174	175	175	190	259	277	252	189	160	167	195	165
Collaborators													
Avg. of Collaborators / Author (CA)	2.2348	2.0109	1.7035	1.7512	1.6031	1.5588	1.5087	1.5935	1.6951	1.8013	1.8947	1.924	2.1485
Max of Collaborator / Author (CA)	373	93	88	93	93	93	111	129	130	125	182	172	214
Avg. of Collaborators / Author (US)	0.6515	0.7105	0.6977	0.6914	0.6145	0.5852	0.569	0.5953	0.6186	0.6429	0.6545	0.644	0.6953
Max of Collaborator / Author (US)	43	174	175	175	54	55	59	66	71	71	79	83	94
Repeated Collaborations													
# of Repeated Collaborations	7310	6911	7165	8033	8172	8155	8750	10861	13405	17340	21055	23423	25525
Avg. of Repeated Collaboration / Author	0.4816	0.4628	0.4528	0.4548	0.4378	0.4179	0.4052	0.4186	0.439	0.4617	0.4689	0.4624	0.4482
Max of Repeated Collaboration / Author	26	36	29	37	39	46	48	43	41	34	35	33	30