Evaluating the Dynamics of Knowledge-Based Network Through Simulation: The Case of Canadian Nanotechnology Industry

Nuha E. Zamzami

Concordia Institute for Information Systems Engineering

Thesis submitted in partial fulfillment of the requirements for the degree of
Master of Applied Science in Quality Systems Engineering at
Concordia University, Montreal, Quebec, Canada

February 2014

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Nuha E. Zamzami

Entitled: Evaluating the Dynamics of Knowledge-Based Network Through Simulation: The Case of Canadian Nanotechnology Industry

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

M.A.Sc. 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. A. Ben Hamza ____________________ Chair

Dr. N. Bouguila ____________________ Internal Examiner

Dr. G. Gopakumar ____________________ External Examiner

Dr. A. Schiffauerova ____________________ Supervisor

Approved by ____

Chair of Department or Graduate Program Director

____

Dean of Faculty

Date ____

Abstract

Evaluating the Dynamics of Knowledge-Based Network Through Simulation: The Case of Canadian Nanotechnology Industry

Nuha E. Zamzami

Collaboration is a major factor in the knowledge and innovation creation in emerging science-driven industries, where the technology is rapidly changing and constantly evolving, such as nanotechnology. The scientific collaborations among individuals and organizations form knowledge co-creation network within which information is shared, innovative ideas are exchanged and new knowledge is generated. Although various simulation attempts have been carried out recently to analyze the performance of such networks at the firm level, the individual level has not been much explored in the literature yet.

The objective of this thesis is to investigate the role of individual scientists and their collaborations in enhancing the knowledge flows, and consequently the scientific production within the Canadian nanotechnology scientists. The methodology involves two main phases. First, in order to understand the collaborative behavior of scientists in the real world, the data on all the nanotechnology journal publications in Canada was extracted from the SCOPUS database and the scientists' research performance and partnership history was analyzed using social network analysis. Moreover, the predominant properties that make a scientist sufficiently attractive to be selected as a research partner were determined using data mining and through a questionnaire sent directly to the researchers selected from our database. In the second phase, an agent-based model using Netlogo has been developed to simulate the knowledge-based network where several factors regarding the ratio, existence and absence of various categories of scientists could be controlled.

It was found that scientists in centralized positions in such network have a considerable positive impact on the knowledge flows, while loyalty and cliquishness negatively affected the knowledge transmission. Star scientists appear to play a substitutive role in the network as most famous and trustable partners to be selected when usual collaborators are scarce or missing. Besides, the changes in the performance of some categories in case of the absence of others have been also observed.

The major contribution of this work stems from the fact that the developed simulation model is the first one, which is fully based on the real data and on the observed behavior of the scientists in knowledge-based network.

Acknowledgements

Foremost, I would like to express my sincere appreciation to my supervisor, Dr. A. Schiffauerova, for giving me the opportunity to become a member of her research group. I would never have been able to finish my thesis without her motivation, patient guidance, understanding and continuous support.

The most important people that I would like to express my deepest gratitude to are my amazing family. I will never ever be able to pay back my dad who has been always the source of my strength and the main reason for every success in my life. The warm words and prayers from my mother have been always lighting my way. I can’t thank enough my elder sister Hanadi for her continues encouragement, endless love and honest wishes. I’m also thankful to my lovely younger sister Rana who considers me as her role model, which makes me keep seeking for the best just not to disappoint her.

I owe my sincere thanks to my warm-hearted friends for believing in me and standing by my side throughout the program. Special thanks go to Alaa Jawa, Basim Al-Badri, Khloud Elaimi, Reham Fadul and Fatimah Al-Zamzami who have been always there for me with all their love, support, and invaluable comments. I have been always confident that I would be able to continue through the rough times as long as I have such true friends to count on.

Last but not least, the financial support from the ministry of higher education in Saudi Arabia is hereby acknowledged and appreciated. This support enabled myself to attend this program and do this extensive research, without which my studies could never have been completed.

Table of Contents

Table of Tables

Table of Figures

1.0 Introduction

In today's rapidly growing technological fields the sources of knowledge are widely distributed. Solving new rising issues and answering many complex and multidisciplinary research questions require higher level of skills and comprehensive knowledge. This leads to the need for collaborative knowledge sharing whose ability to address more complex and critical research problems has already been demonstrated in the literature. Moreover, a significant increase in research productivity as result of scientific collaboration has been suggested by several scholars (Lotka, 1926; de Beaver and de solla Price, 1966; Zuckerman, 1967; Glanzel and Winterhager, 1992; Landry, 1996).

The collaborative activities can be mapped as a complex network, where its nodes represent the collaborators and their partnerships form the links. In such networks, the knowledge is created and transmitted by socially connected individuals whose collaborations shape the links of the networks. In this thesis, it is the individual researchers who are the network nodes and their co-authorship of journal articles is the partnership linkages between these nodes, *i.e.* partners co-creating knowledge through scientific collaboration.

The knowledge creation network is a dynamic social network where the behavior of collaborators is influenced by their interactions with others over an interval of time. Scholars have analyzed such networks in the last decades in order to derive policy implications and to enhance the institutional and governmental decision-making in the area of innovation policy. The existing research studies mainly focused on the firm level analysis, where different categories of firms, for example based on the experience, such as leader and startup, or the organization type such as academic, industry and government are examined (Nagpaul, 2002; Albino *et al.*, 2006; Pyka *et al*., 2009 and others). Nevertheless, there is only one study, which recently examined the role of individual collaborators using simulation approach (Alizadeh and Schiffauerova, 2012). Given the novelty of this research avenue, several research gaps could be recognized.

The purpose of this work is to study the Canadian nanotechnology knowledge-based network at individual level with the focus on the role of scientists and their collaborations in enhancing the knowledge flows and transmission. The mapping of this network is based on the co-authorship relationships between Canadian scientists publishing in the field of nanotechnology. The network is then to be studied in a dynamic context to evaluate its productivity and knowledge flows efficiency.

There are two main research objectives of this thesis. First is to analyze and understand the collaborative knowledge sharing behavior in the real world, and second is to simulate the Canadian nanotechnology knowledge-based network and evaluate its dynamics under several controllable factors. Different research methods are used to accomplish these objectives including social network analysis, data mining, and agent-based modeling approach. The justifications and uses of each method will be discussed in details in Chapter 4.

As the overall productivity of the network depends on the performance of its actors, the quantity and speed of knowledge diffusion is greatly affected by the individual collaborative activities (Pyka and Küppers, 2002). That is, individuals with specific characteristics would facilitate the network's evolution and the behavior of others might discourage it. In the present thesis we aim to identify and characterize scientists who are critically important for the knowledge creation and transmission. The results of this work could thus be used by governmental agencies and other institutions in improving the research and technology polices.

The present thesis is structured as follows: In Chapter 2 the relevant research work regarding the networks of collaborators and the use of the dynamic approach in the study of its performance is reviewed from the literature. The research gaps are identified and the description of research objectives is then presented in Chapter 3. Chapter 4 describes the data and the methodology used for the analysis within six main sections; (a) Scientific production in nanotechnology, (b) Network representation and structural analysis, (c) Data analysis, (d) Survey for influences on partnership decision, (e) Simulation model building, and (f) Experimental scenarios. The analysis and results are then reported in Chapter 5. The main findings are then summarized and discussed in Chapter 6. Lastly, we concluded the remarks and proposed some new research opportunities that this work leads to in Chapter 7.

2.0 Literature Review

This chapter of the thesis reviews the relevant research work from the literature in three main sections. Firstly, the literature on correlation between the collaborative behavior of scientists and the knowledge diffusion including motivations, pros and cons of scientific collaborations has been analyzed. In the second section we have reviewed the studies on how the knowledge-sharing network is created as well as two types of such network are discussed. Finally, the last part reviews the literature about the dynamical approach for analyzing the network of collaborators and compares prior research studies in simulating such networks.

2.1 Innovation Networks and Collaboration

2.1.1 Introduction about Scientific Collaboration

Innovation networks have been conceptualized as a group of socially interacting actors who collaborate in order to obtain, transmit and generate knowledge and present new ideas (Phelps *et al.,* 2012). The main activities these actors perform within the networks are sharing resources and exchanging knowledge through their dynamic interconnections (Meyer, 2003; Powell and Giannella, 2009; Tödtling, 1999) This process aims to enhance the quality and economic value of knowledge which is created (Singh and Fleming, 2009) as well as to increase research productivity (Lotka, 1926; Price and Beaver, 1966; Zuckerman, 1967).

Several researchers have discussed where such behavior may take place. Allen (1983) and Cowan and Jonard (2003), for example, have proposed that, in general, there are three places where collective invention occurs: in nonprofit institutions such as universities, in profit seeking firms and in individuals' minds. Katz and Martin (1997) suggest that individuals, groups, departments, institutions, sectors and countries can form new combinations of ideas. Regarding the industrial sector, Lavie and Drori (2012) stated that collaboration is a major factor for innovation in emerging science-driven industries, and many other researchers (for example Powell and Brantley, 1992; Smith-Doerr *et al.*, 1996; Hagedoorn and Duysters, 2002; Cowan and Jonard, 2003; Soh and Roberts, 2003) highlighted the importance of collaboration in the firms where the technology rapidly changed or evolved. Still, collaboration has been considered as a major factor for innovation in no- and low-tech firms as well (Bross and Zenker, 1998). Powell (1996) also argued that firms in fields with distributed resources, such as biotechnology, showed better innovation performance when they are in networks rather than individual firms.

Several studies on collaboration suggested significant increase in research productivity by the collaborative activity (Lotka, 1926; de Beaver and de solla Price, 1966; Zuckerman, 1967; Glanzel and Winterhager, 1992; Landry *et al*., 1996). Beaver and Rosen (1979), for example, analyzed scientific papers of the French elite in the seventeenth and early eighteenth centuries considering the percentage of articles written by co-authors and showed a positive relationship between collaboration and higher productivity. Similarly, Allen (1983) afterward argued that collaborative knowledge creation played an essential positive role in the innovation performance during the nineteenth century. Drejer and Vinding (2006) based on the data from 441 firms which participated in the CATI (computer-assisted telephone interviewing) survey, found that the frequency of innovative activity measured by the number of patents the firms registered was 1.7 times higher for the firms with regular knowledge sharing processes than for those who do not share their knowledge. Likewise, it has been proved by Manley *et al.* (2009) that the knowledge sharing strategy was 10 times more likely to be used by highly productive firms than by the ones with lower productivity considering their conformity to a list of technological and organizational approaches.

Melin (2000) has conducted a survey and a number of interviews with 195 researchers in Sweden to point out some benefits of collaboration. Most of the answers suggested that the growth in knowledge and the enhanced scientific quality are the most important ones. Other benefits mentioned in the literature were for example, staying aware of competitors' activities if the collaboration takes place between competing firms (Soh and Roberts, 2003) or, in case of start-up firms, benefiting from the expertise accumulated within large organizations (Shan *et al.*, 1994; Stuart, 2000). More examples of benefits are sharing the expenses (Hagedoorn and Schakenraad, 1993; Stuart, 2000; Lavie and Drori, 2012), reducing the risk of failure (Cowan *et al.*, 2007) and integrating the thoughts, proficiencies and assets owned by various firms (Nohria and Garcia Pont, 1991; Gulati, 1995; Stuart, 2000; Melin, 2000; Beaver, 2001; Guimerà *et al.*, 2005).

2.1.2 Motivations on Scientific Collaborations

Several factors are significantly related to the collaboration and/or affecting the individual performance. For example, field is one of the most important aspects in science studies. Collaboration and productivity patterns are influenced by different research cultures and environments owned by different disciplines (Lee and Bozeman, 2005). Indeed, we assume that this factor is even more important in the nontechnology sector where scientists are from diverse engineering, science and medicine specializations.

Other important factors that proved to affect the selection mechanism are the similarity of goals, required skills of the partner, and records of prior satisfactory collaborations (Mat *et al.*, 2009). Moreover, nationalistic motivation in collaboration (collaborators are of same nationality or same cultural background and share the same native language) has shown also significant relation to the collaboration (Lee and Bozeman, 2005).

The responses from 195 university professors to the survey conducted by Melin (2000) about their motives for collaboration showed the significance of 'co-author has special expertise 41%', 'co-author has special data or equipment 20%' and 'social reasons: old friends, past collaboration 16%'. Indeed, several studies discuss that the process of searching for new partners is time-consuming, consequently, scientists most of the times prefer to remain loyal to their previous partners, even when better choices are available. However, it is expected to see some people changing their partners frequently, while others prefer loyalty to their current partners (Buchan *et al*., 2002; Kollock, 1994).

Likewise, Bozeman and Corley (2004) examined the collaboration strategies of 451 scientists and engineers at academic research centers in the United States. The examined aspects were gender, funding and cosmopolitanism[1] as factors to impact the collaboration choices. Their main findings show that females more care about the gender of the partners where those in the study are having 84% of their collaborations

---

[1] Cosmopolitanism is the extent to which scientists collaborate with those around them (one's research group, one's university) as opposed to those more distant in geography or institutional setting (other universities, researchers in industry, researchers in other nations).

with females. Undoubtedly, those with larger grants are being more attractive to be selected as partners and hence have higher number of collaborators. Also, most of them tend to work with the people in their own work group except for industrial applications. Terms such as remote collaboration, distributed collaboration, scientific collaboratories, and international collaboration arise in the literature which means that the geographic location of scientists is also related to the collaborative activities. Other common motivations for assembling a collaborative team include the need to gain access to expensive appliances, unique scientific data, scarce natural and social resources, and large amounts of scientific funding (Guimerà *et al*., 2005; Sonnenwald, 2007).

2.1.3 Pros and Cons of Scientific Collaboration

Sonnenwald (2007) categorized the gains of the collaborative activities into five groups; scientific, political, socio-economic, resource accessibility and social benefits. First of all, the scientific benefits which can be demonstrated by the ability to solve new rising issues and to answer many complex and multidisciplinary research problems through sharing knowledge. In today's technological fields the sources of knowledge are widely distributed and no single firm has all the necessary skills and knowledge to conform the rapid growing technologies (Katz and Martin, 1997; Powell and Brantley, 1992; Hagedoorn and Duysters, 2002). In other words, producing innovations today is no longer possible without comprehensive knowledge and a variety of skills. As a consequence, in order to generate new knowledge and extend the scope of research projects, firms seek the needed skills and expertise through the interaction with partners from dissimilar disciplines (Bozeman and Lee, 2005;

Alvarez, 2012). Such behavior assists in enhancing the research reliability, accuracy and quality by considering these factors from different aspects (Beaver, 2001).

In addition, the national and international political relationships are also influenced by the scientific collaboration. This activity can raise the level of understanding between countries and drive the peace (McGinley and Charnie, 2003). By addressing global scientific problems, nations can satisfy the need in each country while maximizing the profit of available funding (Mervis and Normile, 1998). Other political benefit mentioned in the literature is the scientific and economic expansion that occurs when scientists from advanced and developing counties are collaborating (Velho and Velho, 1996).

Furthermore, there are some socio-economic outcomes from the scientific collaboration. Based on their study of 12 inter-company networks, Wissema and Euser (1991) stated that the collaboration helps firms to gain additional market knowledge or complement each other's knowledge. This kind of benefit is more probably to be gained in the case of business-university collaboration. As addressed by Lambert (2003) in his review report about business-university collaboration submitted to the UK government, this form of collaboration gives firms the opportunity to attract students and scientists for employment purposes. Some countries also establish organizations, such as VINNOVA[2] in Sweden, to support this collaborative activity between academia and industry in order to strengthen the innovation and economic growth (Cohen and Linton, 2003). Hence, sometimes this activity is mandatory for firms and/or universities to be eligible for certain funding or governmental grants.

[2] VINNOVA is a Swedish government agency, was found in 2001, working under the Ministry of Enterprise, Energy and Communications and acts as the national contact agency for the EU Framework Programme for R&D.

Resource accessibility is another benefit of collaboration according to Sonnenwald (2007) categorization. Indeed, scientific collaboration normally provides an access to expensive appliances, unique scientific data and specific social resources owned by the partner (Wagner *et al*., 2002; Wray, 2002; Birnholtz and Bietz, 2003). Accessing the resources also helps to improve the strategic performance for organizations by developing an enhanced competitive advantage (Galaskiewicz and Zaheer, 1999; Gulati *et al*., 2000). According to Phillips *et al.* (2003) the greatest strategic benefit gained from collaboration stems from sharing knowledge, acquiring new distribution outlets, pooling and transferring of all kinds of resources and building a greater understanding of new markets.

After all, the scientific collaboration has a great impact on expanding the social network and developing the personal and professional relationships, which enhance the creativity (Lavie and Drori, 2012). Co-authorship and co-inventorship relations provide a great opportunity to new collaborations by maintaining a strong tie with the partners themself or by establishing a new relation with other individuals in their professional or organizational network.

In fact, the frequency of scientific collaboration has been steadily increasing due to the extensive evidence related to its ability to address complex and critical problems (Sonnenwald, 2007). The growth in the popularity of the research collaboration can be traced through the percentage of co-authored papers as opposed to the ones written by sole author, which has been continuously rising over the past two decades in every scientific discipline, as well as within and across countries and geographic areas (Grossman, 2002; Moody, 2004; Wagner and Leydesdorff, 2005). Grossman (2002) indicated a slight increase in the rate of publication over the last 50 years but with a remarkable increase in the level of collaboration.

On the other hand, some authors argue that collaboration can sometimes pose some challenges that might make it not worth or negatively affect the productivity. The transaction costs associated with working in partnership have been considered as the most critical explanation (Landry and Amara, 1998). Additional costs that might make the collaboration unworthy are the ones incurred in the processes of searching for suitable partners, negotiating and crafting contracts (Williamson, 1983). Other factors are those that lead to a waste of time and energy even in the best collaborative relationships. The examples of such wastes are the time spent on waiting for others to comment, respond, or do their duties. As a consequence, some projects that have been carried out in collaboration were never finished nor had satisfactory results (Bozeman and Lee, 2005). Indeed, excessive collaboration with higher number of collaborators consumes a lot of time and effort needed for managerial issues to maintain relationships and coordinate joint activities (Ocasio, 1997). Lavie and Drori (2012) suggested limiting the collaborative activities especially in case of internal resources availability.

2.2 The Network of Collaborators

Collaboration networks are an emergent phenomenon in Europe and America, and generally involve, not only firms, but also individual scientists, independent research labs, universities, and government agencies. The activities of collaborators and their partners can be mapped with a complex net, where several actors are represented as nodes (or vertices or agents), the linkages represent their collaborative relations and a new knowledge is the product of interplay between them. The nodes have several properties such as gender, age, and affiliation that allow a certain one to be distinguished of others. Agents can function independently or collaboratively and their actions can be specified by several rules or environmental situation. The behavior of the agents in social networks is influenced by their dynamic interaction. Additionally, the essential variables associated with an agent's current situation can be represented by its state (Macal and North, 2010). On the other hand, the links can be *directed* – an *arc*, or *undirected* – an *edge* and represent different relational types and, further, they have weights to show the strength of the relationships (Mali *et al*., 2012).

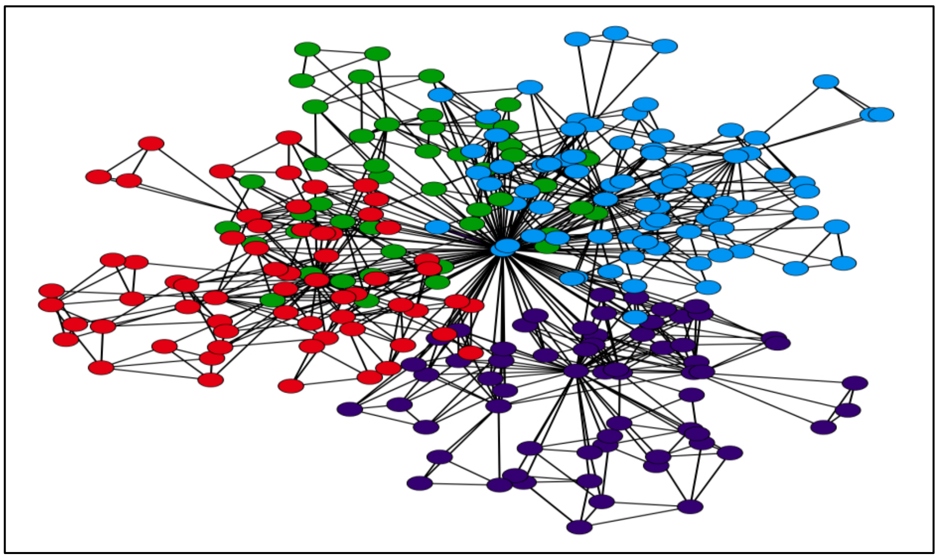

Figure 1: An example of a complex network diagram

There are two kinds of collaborative networks commonly studied in innovation research. First is the network of individuals, which connects all the individual researchers or inventors co-creating knowledge or innovation, and the second one is the networks of firms, which involves cross-organizational collaboration aiming at knowledge creation or application development. The following two sections are dedicated to the review of the research related to each of these networks.

2.2.1 Networks of individuals

The network of individuals can be created based on various kinds of social and collaborative relationships. The Internet (Barabási and Albert, 1999; Broder *et al.*, 2000), Hollywood movie actors' network (Watts and Strogatz, 1998; Amaral *et al.*, 2000; Barabási *et al.*, 2002), and the networks of LINUX and Free software developers (Cowan and Jonard, 2003) are all examples of individual networks. The growth of individuals' networks occurs by increasing the number of either the actors or the links or both (Barabási and Albert, 1999; Barabási *et al.*, 2002).

In case of the innovation networks, it is the individual researchers, or inventors, who are the network nodes. The analysis of co-authorship of research articles and the co-inventorship of patents are the most commonly used methods employed to trace the linkages between these nodes, *i.e.* partners co-creating knowledge or innovation. The connecting link between two scientists in the network is created if at least one paper has been coauthored by them (Newman, 2001a; Barabási *et al.*, 2002; Boccaletti *et al.*, 2006) or if they have co-invented a patent together (Fleming and Frenken, 2007).

By examining the data on articles published in the period of 1995-1999, Newman (2001a) measured the average distance between scientists and several other statistical properties for the collaboration networks that he has created from different databases for biomedical research, physics, high-energy physics and computer science. Newman (2001b) continued his study of the four databases considering multiple numeric factors, such as numbers of authors per paper, the degree of clustering in the network, etc., and indicated different patterns of collaboration for each fields studied.

The similarities and incompatibilities of collaboration patterns have been examined later by Glänzel (2002) for biomedical research, chemistry, and mathematics fields considering all articles with citations counted in the year of publication and the two subsequent years as recorded in the annual volumes of the Science Citation Index (SCI) of the Institute for Scientific Information. The author found that although the co-authorship activities have become considerably more frequent in all disciplines, their impact on the field's productivity and citation rate greatly differs among various research fields. Likewise, Newman (2004) used other three databases in biology, physics and mathematics and concluded that both similarities and contrasts are varying among the fields and changing over the time.

The relation between publishing productivity and collaboration has been examined by Bozeman and Lee (2005) in their research study based on the data from 443 academic scientists. They considered the normal count (number of publications) and the fractional count (number of publications divided by number of co-authors) as measurements for productivity. The authors found a remarkably positive relationship between collaboration and productivity. However, the study did not address any measure of the quality of collaboration and whether it was successfully completed or not. It also did not consider the impact of loyalty, *i.e.* how the repetitive collaboration among the same partners differs from collaborative relations while changing partners.

This limitation has been overcome when Van Segbroeck *et al.* (2009) studied a dynamical graph where they could adjust the behavior and the social ties and observed that the scientists prefer to keep collaborating with the same partner even in the case when an alternative is available. The authors however suggested that being committed to limited social ties could negatively affect the scientific evolution. In contrast, Abbasi and Altmann (2011) later used social network analysis (SNA) measures to show that maintaining a strong tie with a previous partner leads to a better performance than having several co-authorship relationships with multiple ones. Their result was based on a theoretical model according to the reports of five information schools (iSchools) and the citation data was collected from Google Scholar and Association for Computer Machinery (ACM) portal using a web-based application (AcaSoNet[3]). Similarly, Abbasi *et al*., (2012) explored the co-authorship network based on publication data of high impact factor journals in the field of ''Information Science and Library Science'' between 2000 and 2009, extracted from Scopus. Their database contained 4837 publications reflecting the contributions of 8069 authors and they used the structural holes theory to evaluate the network efficiency associated with the scholarly performance (i.e., g-index). The result showed that maintaining a strong co-authorship relationship with one primary co-author led to better performance comparing to having many relationships to the same group of linked co-authors.

Several studies proposed that being in a short geographic distance is not as important as having a social relationship with the partner. For instance, Buchan *et al.* (2002) indicated a considerable decrease in collaboration level as a consequence of an increase in social distance, *e.g.*, lower level of mutual trust or distinct cultural identity. Breschi and Lissoni (2003) later examined three groups of patents applications by Italian firms to the European Patent Office (EPO) in 1987 to 1989 which have been cited at least once by the end of 1996. The outcome of this study proved that in order

[3] A Social Network System for Analyzing Publications Activities of Researchers

to exchange knowledge, social interaction between scientists is a must regardless how physically close they are. However, this study had a limited size sample and did not consider any approach to check the accuracy of the data provided, thus the result cannot be generalized.

Similarly, based on the data of US inventors working in the field of organic chemistry, pharmaceuticals and biotechnology, Breschi and Lissoni (2006) proposed that the social distance between patent inventors is the key factor which influences the productivity, while the role of the geographical distance is much more limited. However, with respect to geographic clusters, Gittelman (2007) considered an average of 1,500 miles as a distance between the co-authors to be considered as spatially clustered. The regression analysis of 5,143 collaboratively authored articles in biotechnology indicated a higher citation in papers that are subsequently cited in the authoring firms' patents co-authored by geographically clustered teams, where the ones co-authored by global teams have higher citation in the scientific literature but less cited in the authoring firms' patents. In contrast, around two-thirds of IT developers see no benefit from locating in the same geographical area than if they are located in geographically distant places (Huber, 2012).

A wide-ranging analysis of high-energy physics dataset consisting of 29,555 papers has been performed by Kas *et al*. (2012). They used several research methodologies such as social networks centrality analysis, topological analysis, investigation of power law characteristics, time series analysis of publication, collaboration frequencies and citation graph. The study aimed to provide a deep understanding of the processes that form the complex co-publication networks and activities of networks expanding by the appearance of new papers and authors in this research field. As expected, the social connections showed a significantly positive impact on publications as authors tend to co-author with those who are within their social networks. Balconi *et al.* (2004) highlighted the critical role of academic inventors in connecting individuals and network components. Generally, academics work in larger teams, exchange information with more people and across more organizations.

Another issue discussed in the literature is the effect of the mobility of researchers, or inventors, who move between institutions within the same geographical region. Knowledge is transportable along with people who master it so their movement from the place where they originally learnt, researched, and delivered their inventions supports knowledge evolution. Breschi and Lissoni (2005) considered a sample of 2,321 Italian patent applications to the European Patent Office (EPO) over the period 1987 to 1989, which have received a total of 5,066 citations by the end of 1996, to discover the role of inventors' movement among companies in promoting the knowledge flows. The authors applied the JTH methodology[4] and indicated a localized expanding in the social network as a consequence of inventors' activity across firms' boundaries.

Fleming and Frenken (2007) later intended to investigate the causes of sudden expansion in Silicon Valley inventor networks in comparison to the one in Boston considering a database of 2,058,823 inventors and 2,862,967 patents registered in 2003 at the United States Patent and Trademark Office (USPTO) and performed the

[4] Jaffe, Trajtenberg and Henderson (1993; hereafter JTH) argued that knowledge spillovers may be measured by the "citations to prior art" contained in most patent documents, and produced a statistical experiment showing that such citations come disproportionately from the same geographical area of the cited patents. The basic JTH methodology has become a classical reference for any empirical work on the geography of innovation.

analysis of the collaboration networks between these inventors through SNA. Moreover, based on the network analysis they identified the key principal actors and conducted interviews with them. The authors illustrated the inter-organizational inventors' network improvement by the labor mobility between established firms. This result can be generalized due the large sample size used and also due to the high level of accuracy of the in the procedure of USPTO data since the USPTO procedure for the registration of innovations involves a rigorous verification to check the information provided by the applicants on inventors.

On the other hand, Breschi and Lissoni (2006) reapplied the JTH experiment on new data of 63,188 US inventors and their 66,349 patent applications at the European Patent Office (EPO), filed between 1978 and 2002 in the fields of organic chemistry, pharmaceuticals, and biotechnology. Breschi and Lissoni (2006) ended up with three inventors' maps representing the network of all inventors in each field and performed the tools of SNA. The result showed that cross-firm inventors, whose name has been reported in patent documents assigned to different organizations, play a critical role in connecting inventors from different firms so they essentially contribute in creating the social network between firms to spread the knowledge.

The international scientific co-operation and its impact on the performance have also been discussed in the literature. Glänzel *et al.*, (1999) studied the role of European Union countries as partners for both advanced and developed nations. The authors analyzed all the papers recorded as article, letter, note or review in the 1985-1995 journal citation reports volumes of the Science Citation Index (SCI)[5] in eight major fields considering the Relative Citation Rate (RCR) as a performance indicator. The results show greater benefit of collaboration for less advanced countries, but also advanced countries benefit from collaboration. Glänzel (2001) later studied the relation between international co-authorship and both national research profiles and citation impact. His study was based on papers published in 1995 and 1996 and citations obtained in 3-year periods for 50 most active countries in all fields. Papers involving international collaboration have been shown to receive higher citation rates compared to purely domestic ones. However, the national citation rates vary considerably among the countries.

Glänzel and Schubert (2001) continued the previous studies concentrated on the RCR of German-Japanese chemistry papers published in 1995 and showed that international co-authorship links displayed a characteristic pattern reflecting geopolitical, historical, linguistic, etc. relations among countries. The authors illustrated that there is no correlation between the strength of co-authorship links and the relative citation reputation of the resulting publication. Afterwards, Glänzel and de Lange (2002) statistically analyzed the citation pattern model they developed earlier in 1997 to study the Multilateral Collaboration Index (MCI) as a function of the share of internationally co-authored papers involving three or more countries. The results of the two previously mentioned studies have been integrated to prove that the number of overall publications is approximately equivalent to the number of international co-relationships.

[5] Science Citation Index (SCI®) is an Institute for Scientific Information (ISI) publishes journals designed primarily as bibliographic tools to help researchers discover and recover publications pertinent to their interests

In spite of all positive results mentioned above, a regular quantitative monitoring of costs and outcomes of international scientific collaboration, *i.e.* bibliometric surveys, must be conducted to insure that it is cost effective (Glänzel and Schubert, 2005). The quantitative effects of such co-operation activities have been studied later by Inzelt *et al.* (2009). Targeting faculties of medicine, sciences, social sciences and humanities in six medium Hungarian universities, a total of 9,585 publications and their citation data for the period 2001–2005 were retrieved from the Web of Science database. The authors compared international and domestic co-authored papers and found a positive effect of international cooperation on citation rate.

2.2.2 Inter-Firm Networks

Studies examining the effect of inter-firm partnerships on firm innovation performance, however, provide inconsistent conclusions. Several studies have stated that the more inter-organizational partners a firm has, the greater probability to it be innovative (e.g., Ahuja, 2000; Owen-Smith and Powell, 2004; Shan *et al*., 1994), and thus many national and international funding agencies and policy makers have been encouraged to get involved in the collaborative R&D arrangements (Benner and Sandström, 2000; Scholz *et al*., 2010). On the other hand, other researchers suggest that the increasing reliance on partnerships for knowledge creation can negatively influence the performance (Rothaermel and Alexandre, 2009; Wadhwa and Kotha, 2006), considering that the costs of seeking a new partner, maintaining the inter-organizational relationships and managing the work among increasing number of partners can exceed their knowledge-creating benefits.

Wissema and Euser (1991) studied 12 outstanding Dutch innovation networks for new technological inventions authorized by and executed in collaboration with the NEHEM[6]. Their aims were to investigate the reasons for companies' collaboration especially in this area. They highlighted the different types of technological innovation collaboration and pointed out some factors for successful innovation networks. The correlation between the innovation performance of startup firms and their collaboration with established ones has been investigated later by Shan *et al.* (1994). They analyzed the performance of 85 startup firms in biotechnology industry who were in collaborative agreements before 1989 and concluded that the inter-organizational scientific relationships positively affect the firm's efficiency.

Mowery *et al*., (1996) analyzed the change in the firm's technological capabilities in the case of resources overlapping with considering the citation pattern of their patent portfolios as an indicator. Based on data of 5000 firms who were involved in over 9000 strategic alliances presented in the Cooperative Agreements and Technology Indicators (CATI) database, the authors suggest that it is more effective for the firms to have an equity joint venture[7] agreement than having contract alliances such as licensing agreements. The analysis of the data also shows that it is significant for the firm to have the ability to recognize the value of information that can acquit through alliances and this ability depends on the pre-alliance relationship between the two firms' patent portfolios. Another result of this study was that the level of knowledge transfer between firms in the same geographical cluster is higher than the one between firms in different countries, which support the argument made by Gulati *et al.* (2000) regarding some

[6] The Nederlandse Herstructurerings-Maatschappij (Dutch Restructuring Company)
[7] The joint venture is a finite time business agreement

restrictions to inter-firm knowledge transfer such as distance, cultural differences, and others. Indeed, recognizing and acquiring the required knowledge for the invention process in technology-intensive firms, such as biotechnology, is considerably complex. Hence, the knowledge-access benefits of establishing and maintaining alliances between these firms can be gained in both formal and informal as well as both in proximate and distance collaborative relationships (Zaheer and George, 2004).

Another issue that has been discussed in the literature is how the inter-firms collaboration agreements influence the performance of the developing firms in particular. By studying life histories of 142 startup firms in biotechnology that began in Canada between 1991 and 1996, Silverman *et al.* (2000) supported their suggestion that due the lack of their resources, the early performance of startups can be enhanced by establishing alliances. The firm's performance has been evaluated considering multiple measures such as revenue growth, employment growth, R&D spending growth, and patenting success. Correspondingly, Soh and Roberts (2003) investigated the growth of emerging innovations as a result of complex technologies development through the expansion of innovators networks between the firms. Based on 150 firms and 319 alliances in the US data communication industry from 1985 to 1996, the authors argue that new complex technological firms are most likely to survive if they are involved in scientific partnerships.

The recombination of knowledge held by the partners prior to the collaboration, and the history of their collaboration play a significant role in innovativeness development. Cowan *et al*., (2007) argued that the probability of successful innovation increases in the case of prior cooperation experience. Hence, the strongly structured innovation network, which could be achieved by collaborating with a limited number of partners, aims to reduce the level of failure. In contrast, based on examining the patent performance of 1,106 firms in 11 industry-level alliances networks, Schilling and Phelps (2007) found an evidence for their suggestion about the positive impact on the firm's performance by having a wide range of collaboration paths, which provide them with more resource accessibility. On the other hand, a survey for a sample of 284 cross-sectorial firms showed that having a large business network does not really impact the innovation performance, unless a formal scientific interactions with clear innovativeness objectives are set (Cantner *et al*., 2010).

2.3 The Simulation Approach

2.3.1 Modeling and Simulation

Computer model is basically a representation of a real or theoretical system that shows its behavior based on some information. By simulating the system, some experiments are performed to observe the changes in the outputs under different conditions and variety of inputs (Ali and Moulin, 2005). A typical dictionary definition from the Oxford English Dictionary describes computer simulation as « *the technique of imitating, on digital computer, the behavior of some situation or system (economic, mechanical, etc.) by means of analogous models, situation, or apparatus, either to gain information more conveniently or to train personnel.* ». Fishwick (1995), a specialist in computer simulation, has another definition for computational simulation as: « *A computer simulation or computer model is a computer program which attempts to simulate an abstract model of a particular system. Computer simulations have become a useful part of modeling many natural systems in physics, chemistry, and biology,* *human systems in economics, and social sciences and in the process of engineering new technology, to gain insight into the operation of those systems* ».

Computer simulation has been primarily recognized as a crucial tool for analyzing complex systems (Banks, 1998; Hao *et al.*, 2006) or at least to validate the analysis when other analytical tools can be used. Simulation is the most accurate manner to describe what is actually happening in the real world (Bonabeau, 2002). It is usually used to understand the structure of the system and the behavior of the nodes in large-scale networks under a variety of conditions (Fujimoto *et al.*, 2003), as well as to contribute in predicting the influence of various assumptions and initial conditions to the current behavior (Axelrod, 1997 a; Pyka and Küppers, 2002; Hao *et al.*, 2006). Modeling and simulation are needed tools to gain a deeper understanding of dynamic systems[8] or to provide a root for managerial decision making regarding the system control or transform in order to improve its performance (Bonabeau, 2002; Glahn and Ruth, 2003; Birta and Arbez, 2007).

Some considerable reasons for using computational simulation as a problem-solving technique have been stated by Gaines (1979) and other researchers in the literature. For instance, models performed to replace the real experiments are usually executed so slowly and consume time, or so fast and hard to be observed, thus simulation aids to control the time of experiments as needed. The expenses of necessary measurement tools, upgrading the hardware and/or communication between nodes to collect the required data have been also recognized as a good justification for the effectiveness of using computerized simulations. Moreover, controlling the variables, and/or accessing

[8] A dynamic system is a collection of interacting entities that produces some form of behavior that can be observed over an interval of time.

system parameters are also good reasons for simulating the systems, whereas, in the real ones, researchers cannot manipulate some inaccessible inputs.

2.3.2 Collaboration Networks Simulation

Nagpaul (2002) examined the pattern of mutual cooperation among fifty elite institutions, the most productive institutions in India, which account for approximately two thirds of all articles contributed by India to the typical journals covered by Science Citation Index. His study mainly focused on cooperation links of an institution with other institutions and on whether the other institution is in the same or different research fields. The data on cooperation relations between elite institutions were taken from the database created for the project "Science beyond Institutional Boundaries", sponsored by the Department of Scientific & Industrial Research (DSIR), Government of India (Nagpaul, 1997). The database, containing more than 50,000 articles, is derived from the CD-ROM version of Science Citation Index for five consecutive years 1990-1994. The articles were categorized into eleven non-overlapping fields: Mathematics, Physics, Chemistry, Biology, Earth and Space Science, Agriculture, Clinical Medicine, Biomedicine, Engineering and Technology, Materials Science and Computer Science. Using block model analysis, Nagpaul (2002) represented how institutes are embedded in a network and being clustered in subgroups "blocks" and then illustrated the relationships among the subgroups. The study result shows that the geographical and thematic proximities of the institutions significantly influence the structure of the network.

Albino *et al.* (2003) proposed a computational approach with a dynamic behavior to study the multiple forms of cooperative and competitive relationships within Industrial

Districts (IDs)[9]. They developed a computational model and performed a simulation analysis to prove the benefits of cooperation for the IDs in balancing the utilization of supplier production capacity and minimizing the customer unsatisfied demand, as well as to evaluate those benefits in different competitive scenarios and diverse ID organizational structures. The results show that the cooperation has a significantly positive impact on the ID performances; however, IDs in different organizational structure perform differently specially in a competitive environment.

An agent based simulation model is later used by Albino *et al*. (2006) to investigate the significant modifications that should be implemented on the current innovation processes in IDs in order to assist their survival in a highly competitive environment. The social network resulting from knowledge flows between the IDs is one of the different types of network among agents and among agents and the environment that have been studied in this agent-based model. Analyzing the system behavior have been achieved by considering 28 agents, where 8 are final firms, 16 suppliers, and 4 infrastructure suppliers. Simulation plan involves different experiments to be run for four different scenarios. Albino *et al*. (2006) considered the number of firms, average firm's resource availability, and standard deviation of firm's resource availability as performance measurements for IDs. The study indicates the significant of using external sources of knowledge and R&D investments in the IDs survival among the competitors. Moreover, it shows that the level of ID innovativeness seems to increase by the existence of the leader within the network.

---

[9] IDs are geographically defined production systems, characterized by a large number of small and medium-sized firms that are involved at various phases in the production of a homogeneous product family

Gilbert and Troitzsch (1999) developed a simulation platform in The Self-organizing Innovation Networks (SEIN) project to investigate the structure and dynamics of technological collaborations using computational experiments (Pyka and Küppers, 2002). The data used to test the model has been gathered on four case studies of innovation networks which are biotechnology in France (Pyka and Saviotti, 2000), knowledge Intensive Business Systems (KIBS) in England and the Netherlands (Windrum, 2000), personal and mobile communications in the UK (Vaux, 2000), and combined heat and power (CHP) technology networks in The Netherlands, Germany and the UK (Weber, 2000).

The general simulation model (GenSEIN) developed by Gilbert and Troitzsch (1999) is a multi-agent system where the actors are represented by an agent or 'object' and those agents are designed to have several intelligent attributes such as autonomy, ability to interact with other agents; reactivity to signals from the environment; and proactivity to engage in goal-directed behavior (Wooldridge and Jennings, 1995). The main objective of applying the model to the previously mentioned case studies is gaining a further understanding of how the networks are evolving, the similarities and differences between them, the effect on their productivity and exploring the factors that encourage the formation of network and the obstacles that might discourage its formation (Pyka *et al*., 2003). Pyka and Küppers (2002) summarized the results of applying the model, which shows that the overall performance of the network depends on the performance of the best actors, who have the highest volume of publications, and who thereby assist in improving the average number of co-authored papers in the whole population. Moreover, the model identifies some factors that characterize the partnership relation as being short-term or long-term by investigating the amount of effort they put into finding suitable partners in different sectors.

Pyka *et al*. (2004) later developed a multi-agent model for Simulating Knowledge Dynamics in Innovation Networks (SKIN) containing heterogeneous innovative firms in a complex environment. The firms in SKIN model are trying to sell their innovations to other agents and end users and they also have to buy raw materials or more sophisticated inputs from other agents (or material suppliers) in order to produce their outputs. The SKIN is modeling the market and the firms' behavior in exchanging knowledge, cooperating and networking with others in order to improve their innovation performance and sales. Each firm in the model has an individual knowledge base, a stock of initial capital and different firm size; however, they all aim to utilize their knowledge for creating innovative products. Partnership is one of the strategies that firms follow in order to learn from other agents, and they make the decision about whether to form or reject cooperative arrangements based on mutual observations.

SKIN allows the investigation of different industries where different strategies have an impact on the firms' productivity with altering several parameters and describing an industry's cooperative behavior. By running some simulation experiments a researcher can find the critical parameters that change the model's results so that they follow the historical sequence of another industry. Several experiments have been conducted on SKIN by Pyka *et al*. (2007) trying to illustrate the impact of different learning activities and emphasize the significance of innovation and learning. These simulation experiments were only the start that shows the possibility of investigating the complex relationships between firm and sector success and organizational learning; through carrying out experiments on a model that would be impossible to perform in the real world.

Several European universities later formed a partnership to conduct a three-year project (2006-2009) Network Models, governance and R&D collaboration networks (NEMO) which aimed to investigate the impact of sets of political governance rules, structures and functions of R&D on the emerging collaboration networks structures in the European Framework. Scholz *et al*. (2010) updated the agent-based model SKIN and introduced a modified version called SKEIN, (Simulating Knowledge dynamics in EU-funded Innovation Networks) intended to simulate the emergence of collaboration networks and knowledge production in EU-funded R&D collaboration projects. The model requires agents, such as university and research institutes departments, and research divisions of firms, to form partnerships in order to be eligible for funding. Experiments on SKEIN model show that both network structure and research productivity are greatly influenced by the political rules, which result in a strongly connected network. The model promotes the emergence of European Knowledge Society by supporting policymakers in their planning to improve the effectiveness and efficiency of R&D funding by considering some polices, such as supporting large projects, encouraging key players and geographical dispersion.

The University-Industry Relationships (UIRs) and their impact on the innovation performance have been analyzed by Triulzi *et al.* (2011) using an agent-based modeling approach. The model focuses on the knowledge dynamics in the biotech and pharmaceuticals sectors considering universities, large diversified and dedicated firms, as well as some research agencies for composing the model’s population. The authors concluded that this kind of relationships aims to turn the research activities from basic to practical approach. Universities in such relationship are benefiting from industry financial resources, while companies are more likely increasing their innovative capabilities. On the other hand, UIRs negatively affect university research orientation that can be mitigated by changing government research policies to more support academic basic research.

Similarly, Ahrweiler *et al.* (2011a) applied the SKIN model to investigate the role of universities in improving the efficiency of innovation network in knowledge-intensive industries. The experiments compared the productivity of the innovation networks in two scenarios; with procedures relying on theoretical frameworks coming from academic partners and in their absence. Result show that the quantity and speed of innovation diffusion have been greatly increased by having universities in the cooperating population. Moreover, the existence of academic partners has a significant impact on increasing the variety of knowledge among the firms and making them more attractive for new collaboration. Ahrweiler *et al*. (2011b) later intended to evaluate the innovation performance by conducting more experiments with different actor strategies and different access conditions to capital. The model analysis suggests that strategic collaborations allow actors to compensate for structural limitations and act effectively within innovation processes. However, the analysis could not provide an adequate explanation for understanding the complex interplay between governance strategies in distributing the public funding and institutional framework. This limitation led to a further work that has been proposed in an undertaken project (2011-2016) called IPSE (Innovation Policy Simulation for the Smart Economy). The project aims to implement and test innovation policy scenarios in the Irish ecosystem using agent-based modeling (ABM) building on SKIN in order to investigate the impact of certain innovation policy strategies on the knowledge dynamics in university-industry-government networks.

European institutions from both academic and industry are currently conducting a research project (2010-2015) called Management of Emerging Technologies for Economic Impact (MANETEI) to examine the emerging phenomena in Irish nanotechnology innovation network. Social network analysis (SNA) tools have been used to evaluate the existing nanotechnology innovation network considering its structure, dynamics, collaboration patterns and interdependencies. The following step is to design an ABM based on the already existing ABM setting SKIN (Pyka *et al*., 2004). The objectives of this project include, but are not limited to, identifying strategies and capabilities needed for different members of technology innovation systems, investigating the non-technical factors that influence the development of emergent technologies, monitoring the impact of emergent technologies, and guiding practitioners in managing emergent technologies (Schrempf, 2013).

Alizadeh, and Schiffauerova (2012) developed an agent-based model simulating the Canadian biotechnology innovation network targeting to examine the impact of individuals' behavior on the network performance. The experiments proposed a significantly positive role for both star scientists and gatekeepers in increasing the efficiency of the network. However, the loyalty of the relationships among the scientists seems to have a negative impact on both performance and knowledge flows.

3.0 Thesis Contributions and Objectives

3.1 Research Gap and Scientific Contributions

The following points address some research gaps and describing the main contributions of this work:

- **The dynamic analysis at the individual-level:** Although various simulation attempts have been carried out recently to analyze the performance of the innovation networks at the firm level (Nagpaul, 2002; Albino *et al.*, 2006; Pyka *et al*., 2007 and others), the individual level has not been much explored in the literature yet. This research aims to study the impact of individuals' collaborative behavior on the overall network productivity and efficiency through developing an agent-based model.
- **An analysis of the correlation between research performance indicators and the social network analysis measurements:** The correlation between scientists' network positions and their research performance has been examined in the literature with limited amount of data and considering only one performance indicator (i.e. for example Abbasi, & Altmann, 2011 and Graf, 2011), while the data in this study consists of wide range of articles published during over 20 years and with various measurements for the authors' research productivity such as number of publications, citations count and H-index.
- **The roles of individual scientists:** The only work that explored the dynamics of innovation networks at individual level is Alizadeh and Schiffauerova (2012) who studied the role of the star scientists and gatekeepers in the innovation

network, as well as the impact of loyalty based on the link age. In this thesis this idea is developed and the scope greatly extended. New roles are to be introduced, and the loyalty will be identified for several groups of scientists characterized by their position in the network, and the role of each group of scientists will be examined.

Moreover, Alizadeh and Schiffauerova (2012) examined the role of gatekeepers and star scientists through two basic scenarios, i.e. the case of their existence and absence, where how their number affect the structure and performance of the network has not been discussed in the literature up to now.

- **The correspondence between reality and conceptual model:** None of the previous literature considered real data when building their models. However, this work is based on the observation of the collaboration patterns from the co-authorship of articles in SCOPUS. The only other work, which studied the co-authorship to build the collaboration behavior into the simulation model (Alizadeh and Schiffauerova, 2012) still made relatively simple assumptions on the individual behavior of scientists. This thesis has incorporated extensive analysis and observations of the behavior of scientists in order to build the partner selection mechanism to the model.

3.2 Research Objectives

The present thesis has two main objectives as described below:

Objective 1: Understand the collaborative behavior of nontechnology scientists in the real world

- Visualize and mathematically analyze the structure of Canadian nanotechnology knowledge-based network
- Determine the key properties that make a scientist sufficiently attractive to be selected as a research partner
- Identify and characterize scientists who are critically important for the knowledge creation and transmission
- Detect the patterns of the research performance and collaborative knowledge sharing behavior for different groups of scientists

Objective 2: Investigate the role of individual scientists and their collaborations in enhancing the knowledge flows within scientists

- Develop an agent-based model simulating the Canadian nanotechnology knowledge-based network using the most appropriate software package
- Run several modeling scenarios where factors regarding the ratio, existence and absence of various categories of scientists change
- Analyze the network productivity and compare the scientific production indicators under different tested scenarios
- Analyze the knowledge transmission efficiency and compare the network structure measurements under different tested scenarios
- Examine the changes in the performance of each group of scientists in case of the absence of others.

4.0 Methodology

The scientific collaborations among individuals and organizations form knowledge creation network within which information is shared, innovative ideas are exchanged and new knowledge is generated. In this thesis we consider the co-authorship relationships among individual scientists as the main components of the innovation network. In such network, the knowledge is created and transmitted by socially connected scientists whose collaborations shape the links of the networks. The nodes of the network represent the scientists, while their collaborations are the links connecting these nodes. Although there are other forms of collaborations between scientists taking place for various purposes, our main focus is on these co-authorship links, because they are the means of the knowledge transmission in the network. They create a complex net of knowledge-based relationships and thereby greatly contribute to the production of scientific publications.

This study is based on real data, which involves all the journal articles in nanotechnology field published within 1980-2012 where at least one of the coauthors is affiliated to a Canadian institution. The present thesis consists of two main stages; first is to analyze the collected data so we can understand the behavior of scientists in the real world and second is to simulate the system where several factors can be controlled. We have created the network based on the co-authorship relationships between the scientists and various network properties were calculated for studying the structure of the network. Based on the social network analysis, we defined five separate profiles that characterize scientists to study their performance. Research activities information about each co-author, such as his/her publications count, citation count and h-index, along with their collaboration history are then used as inputs for data mining procedure to detect the performance patterns and collaboration behavior for each group of scientists. As a complementary approach we ran a survey sent to active researchers having scientific collaborations in SCOPUS, which allowed us to determine the predominant properties that make scientists with different profiles attractive to be selected as research partners. All of this data is then fed into an agent-based model that simulates the collaboration activities of Canadian nanotechnology scientists. Finally, we run out several scenarios with changed settings regarding the ratio, existence and absence of each group. The detailed methodologies used in the present thesis are discussed in the following sections.

4.1 Scientific Production in Nanotechnology

4.1.1 Introduction to Canadian Nanotechnology Industry

As commonly acknowledged, the concept of nanotechnology was introduced by Richard Feynman in 1959, however, the actual term "nanotechnology" was not invented until 1974 by Norio Taniguchi. From the scientific point of view, "Nanotechnology can be defined as referring to materials and systems with structures and components exhibiting novel and significantly improved physical, chemical and biological properties, as well as to the phenomena and processes enabled by the ability to control the material properties on the nano-scale size" (Miyazaki and Islam, 2007).

Nanotechnology is an emerging technology, which has various potential applications that might also have an effect on other scientific disciplines such as advanced materials, biotechnology and pharmacy, electronics, scientific tools and industrial manufacturing processes. Over the last decades, nanotechnology has attracted so many scientists and researchers to get involved in the relevant research in both academic and industry to accomplish more and more of its expected benefits (Hullmann and Meyer, 2003; Miyazaki and Islam, 2007).

Nanotechnology innovation system has been identified as a dynamic process, involving multiple interacting and co-operating actors, variations of essential technologies, society and business models (Carlsson *et al*., 2002). Nanotechnology is very multidisciplinary field, which covers a wide range of nanotechnology disciplines, materials and systems. Meanwhile, there is no formal categorization in the databases of scientific articles. For these reasons, some sets of specialized keywords have been used by the scholars to distinguish the nano-related articles e.g. (Fitzgibbons and McNiven, 2006; Zucker and Darby, 2005; Porter *et al*., 2008). The used combined collection of keywords has been created based on seven different sources and was then consulted with nanotechnology experts (Moazami, 2012).

4.1.2 SCOPUS Database

In order to create the network an extensive data about the individuals, their research performance and their collaborations was needed. The main approach of the thesis consists of the exploitation of the large amount of information extracted from, SCOPUS[10], which is the largest abstract and citation database of peer reviewed research literature. The required data is extracted by Moazami (2012), a member of our research group, from the database using an automated extraction program. Moazami (2012)

[10] **Scopus** is a bibliographic database containing abstracts and citations for academic journal articles. It covers nearly 21,000 titles from over 5,000 publishers, of which 20,000 are peer-reviewed journals in the scientific, technical, medical, and social sciences (including arts and humanities). It is owned by Elsevier and is available online by subscription. http://www.scopus.com

intended to select the most reliable and comprehensive source of data in terms of the diversity of fields, authors' and address information, and number of articles can be retrieved. Based on the comparison of different digital libraries and online databases he found SCOPUS is the most suitable for the research purpose.

The complete database contains around 748,251 nanotechnology articles, where for each of them we have its title, abstract, keywords, references, the information on the publication and the journal and the citation each article received each year. Moreover, for each of the co-authors we have their first and last names, a complete history of their affiliations, total numbers of articles, co-authors, H-index and citations count. United States followed by China and Japan have the largest number of publications overall. The chart below shows the total number of publications in nanotechnology in several countries.

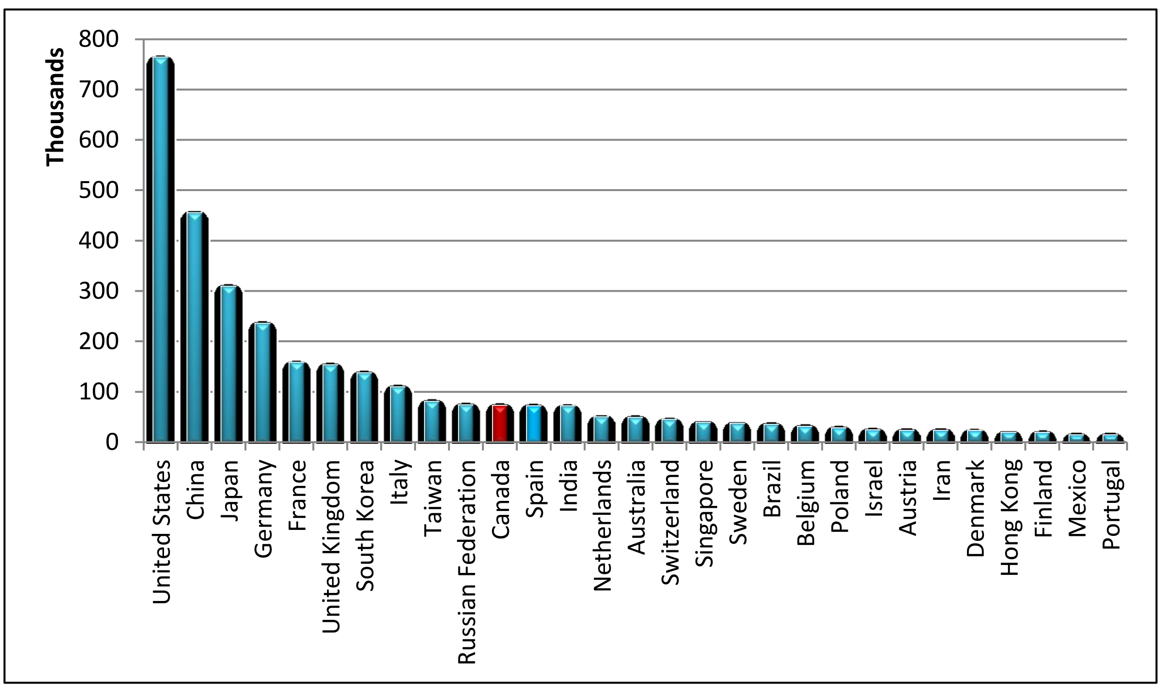

Figure 2: Overall volumes of Scientific output related to nanotechnology

According the previous chart, Canada has been considered one of the most productive countries generating scientific nanotechnology research. Although it is not one of the top leading countries in the field, there is a rapid increase in the number of its publications since the early 1990s. The major concentration of both research and business in Canadian nanotechnology industry can be found in Ontario, Quebec, British Columbia and Alberta. Most of these provinces have organizations that already established or planning to establish province-wide associations to encourage economic development through nanotechnology research. Currently, there are between 50 to 200 companies engaged in nanotechnology-related businesses and a number of research institutes opened by highly ranking universities such as University of Alberta, University of British Columbia and University of Waterloo (Kuroiwa, 2006).

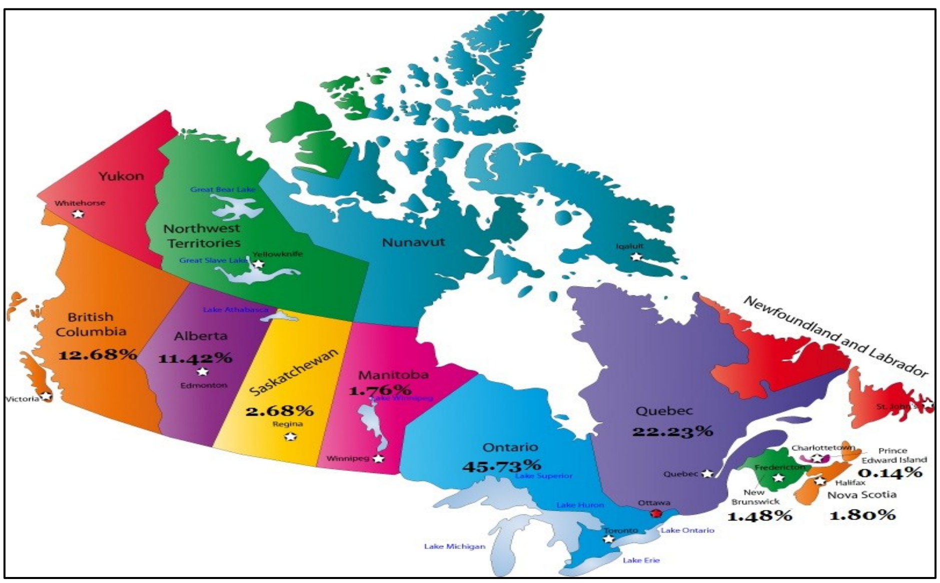

Figure 3: The percentage of nanotechnology authors in each Canadian province

As our study concerns only Canada we have extracted 81,727articles where at least one of the coauthors is affiliated to a Canadian institution. The total number of coauthors is 21,498 including those from outside Canada who are collaborating with Canadian scientists.

Year	Total Publications	Change rate	Year	Total Publications	Change rate
1995	1190	1.25	**2004**	2997	2.26
1996	1293	1.34	**2005**	3894	2.56
1997	1548	1.54	**2006**	5020	2.85
1998	1577	1.56	**2007**	5292	2.91
1999	1648	1.60	**2008**	6897	3.21
2000	1621	1.59	**2009**	6999	3.22
2001	1664	1.61	**2010**	8470	3.43
2002	2096	1.87	**2011**	9516	3.56
2003	2468	2.05	**2012**	12570	3.88

Table 1: The number of nanotechnology publications in Canada and the yearly increase rate

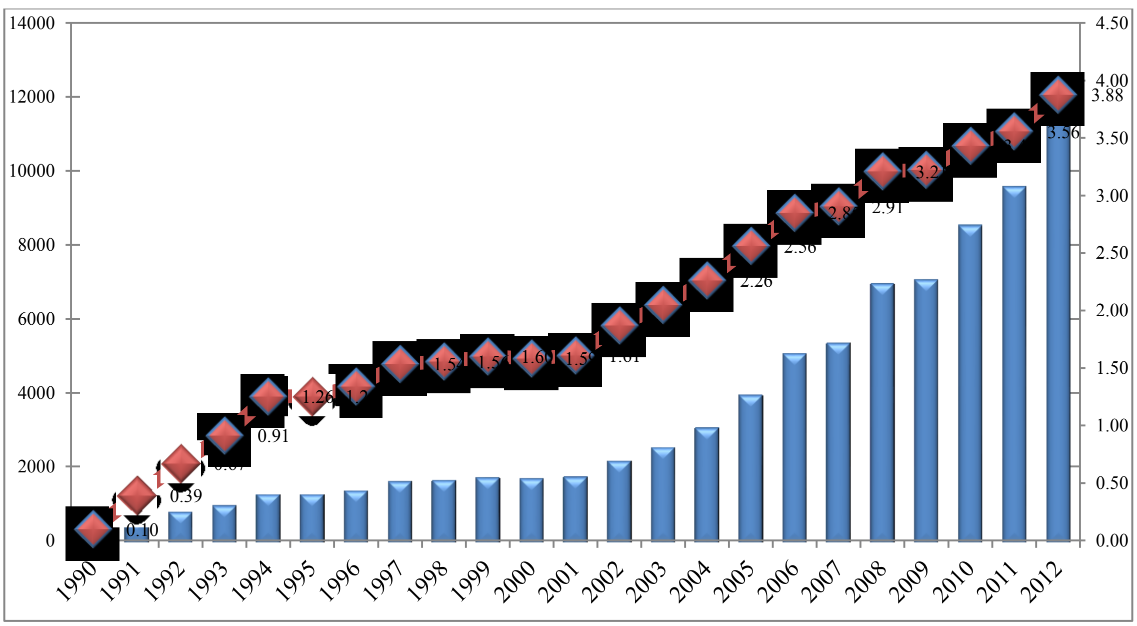

Figure 4: The increase in nanotechnology publications in Canada

The table and figure above showed a major increase in nanotechnology publications in Canada after 1996 with a total number of publications per year exceeding 12,000 in 2012.

4.1.3 Research Performance Indicators

To assess the performance of the researchers in our database we have considered some quantitative and qualitative bibliometric measures that mentioned in the literature:

- **Number of publications**

The total number of publications of each researcher has been calculated based on our database, which means only his/her journal articles in SCOPUS that contain one or more of the specialized keywords in nanotechnology and that were published between 1980 and 2012 have been taken into consideration.

- **Citation Count**

Typically, the higher citations frequency that a researcher's receives for his articles helps him to get a higher visibility and impact in the research community (Lehmann *et al*., 2006; Yan, and Ding, 2011). We have considered the total number of citations, including self-citation[11], for each scholar who appears in our database as an indicator for his performance in terms of the quality of his publications. The citation count information has been collected from SCOPUS as well.

- **H-index**

The h-Index is also being used by SCOPUS and many other academic databases to measure the performance of scholars. It is a simple measure introduced by Hirsch (2005) that combines in a simple way the quantity of publications (i.e., number of published articles) and the quality of publications (i.e., frequency of citations). The h-Index is defined as follows: "A scientist has an h-Index of h, if h of her Np papers have at least h citations each, and the other (Np − h) papers have at most h citations each".

[11] Self-citations are bibliographic references by authors to one or more of their previous publications and constitute a principal basis for objections to the use of citation frequency data as a measure of research quality or impact.

That is, a researcher has an index of h if he has published h papers, which have been cited by others at least h times.

The following table compares between the three measures that used in this thesis for evaluating the performance of researchers based on bibliometric data.

Measures	Advantages	Disadvantages
Number of papers	-Measure quantity	-Does not measure impact of papers
Number of citations	-Measure impact	-Might be overestimated through a small number of highly cited papers with many coauthors. -Gives weight to highly cited articles instead of original research contributions
H-index	-Measure the quantity and broad impact of a research -Eliminates the disadvantages of previous mentioned measures -Deemphasizes single, successful publications	-It is limited by the number of publications -Has less accuracy than the simpler measure -Depends on the person's scientific age and does not account for the number of authors -Never decreases and does not differentiate between active and inactive researchers

Table 2: Summary of advantages and disadvantages of different researchers productivity measures.

The disadvantages of theses measurements have been overcame by new indices proposed by (Abbasi *et al*., 2010) such as Researcher Productivity Index (RP-Index)[12] and Community Productivity Index (CP-Index)[13]. These newly defined indices could be considered for future studies of this work.

4.2 Network Representation and Structural Analysis

4.2.1 Social Network Analysis (SNA)

The social networks analysis (SNA) was introduced in the early 1920s with focus on relationships among social entities, as communication between members of a group, trades among nations, or economic transactions between corporations (Boccaletti, 2006). Social network analysis is a diagnostic method for studying the mechanisms of communication and collaboration between members in different groups (Racherla, and Hu, 2010). By applying it into a particular domain, SNA allows us to identify interaction patterns among network members, the number and structure of the sub-groups within the networks, and their organization and evolution (Anklam, 2003). Some objectives for social networks analysis that are mentioned in the literature include the detection of both strengths and weaknesses within and among research organizations, businesses, and nations as well as the contribution to the scientific development and funding policies (Owen-Smith *et al*., 2002; Sonnenwald, 2007).

Scientific collaborations are defined as ‘‘interactions taking place within a social

---

[12] The basis for the Researcher Productivity Index (RP-Index) is the normalized number of paper citations of a researcher j (NCji). The NCji is calculated as the number of citations of paper i of researcher j divided by the number of years that the paper is available and multiplied by a factor Cji, which represents the contribution of the researcher j to the paper i.

[13] The Community Productivity Index (CP-Index) of a research community k is defined as the largest natural number y such that the top y researchers of this research community have at least in average a value of y for their RP-Index, given that the researchers are sorted according to their RP-Index in decreasing order.

context among two or more scientists that facilitates the sharing of meaning and completion of tasks with respect to a mutually shared, super-ordinated goal'' (Sonnenwald, 2007). According to that definition, those collaborations commonly occur through formal and informal social relationships between individuals across disciplinary, organizational, and national boundaries (Barabasi *et al*., 2002; Sonnenwald, 2007).

Network studies have attracted many scholars in the recent years. Information scientists examined several forms of social interaction networks such as publication, citation and co-citation networks, collaboration structures and others. The value of analyzing social networks consists in its ability to assist with understanding of how to share professional and scientific knowledge efficiently and with evaluating the performance of individuals, groups, or the entire social network. For instance, we can indicate the collaboration activity of a researcher using his social network within a research community (Abbasi *et al*., 2010).

Social network analysis is sometimes also called 'structural analysis' since it aims to understand the social phenomena concerning the relational links. Unlike the traditional individualistic social theory, SNA considers the relationships between actors as first priority and individual behavior as second (Otte and Rousseau, 2002). Structural regularities and how they influence actors' behavior is another main aspect of SNA. Wetherell, *et al*. (1994) described SNA as follows: "Most broadly, social network analysis (1) conceptualizes social structure as a network with ties connecting members and channeling resources, (2) focuses on the characteristics of ties rather than on the characteristics of the individual members, and (3) views communities as 'personal communities', that is, as networks of individual relations that people foster, maintain, and use in the course of their daily lives."

Social networks are represented as a graph constructed of nodes (actors or vertices) and links (ties, relations, or edges). Nodes, which indicate individuals, organizations, or information, are connected with a link, if one or more specific types of relationships (e.g., financial exchange, friendship, trade, and Web links) exist between them. Co-authorship network represents an example of a social network by mapping the graph including authors who have coauthored common publications (Yin *et al*., 2006; Racherla, and Hu, 2010; Staudt, C., 2011).

In this thesis, a node represents a researcher, while a link between two nodes indicates that these two scientists have at least one joined publication. By calculating social network analysis (SNA) measures and several researcher productivity indicators (number of publications, citation count and h-index), we aim to find the correlation between the position of a researcher within the collaboration (co-authorship) network, and his/her research performance. In addition, we will categorize the scientists in different groups in order to evaluate the research performance as well as to detect a pattern of the collaboration behavior of the scientists belonging to each group.

4.2.2 The Social Network Matrix and Properties

A visual representation of social networks provides a deep understanding of large and complex communities such as academic researcher groups (Racherla, and Hu, 2010). Several computer software programs are used as tools for analyzing such networks numerically and visually. One of these effective tools is Pajek[14], pronounced `Payek', which means `spider' in Slovenian. It was specially designed to manipulate, handle and analyze very large networks having on the order of 10^3 to 10^6 nodes. It has been used in academic publications for several years due its flexibility and powerful graphical user interface (GUI) that enables the management of multiple networks, components and analysis outputs at once (Berryman and Angus, 2010).

In order to prepare the data to be analyzed, we needed to create the Pajek format network files. We first extracted the collaboration relationships form our database to a two-column Excel format file. The first column contains the authors' identification number (author id) as it appears in the database, where the second column contains the ids of his/her coauthors. It is possible to have the same row more than one time in case of multiple articles have been coauthored by this pair of scientists. The number of repetition is considered as a weight for the link and represents the frequency of collaboration between two scholars. In addition, reciprocal rows can appear in the dataset by having the scholar as an author once and as a coauthor another time. For network visualization, it is worth mentioning that Pajek ignores non-identical pairs and people who name each other reciprocally unless you specifically instruct it otherwise.

[14] Pajek is developed by V. Batagelj and A. Mrvar, Department of Mathematics, Faculty of mathematics and physics, University of Ljubljana, in 1999. It is freely available for noncommercial use and can be downloaded from the following webpage: http://pajek.imfm.si/doku.php?id=download

The Excel file later has been converted into a one-mode[15] undirected[16] network (.net) format, that is readable by Pajek using the Excel2Pajek[17] tool.

After preparing the social network matrix we have analyzed the network mathematically and calculated the following measures for each node (representing an author in our network).

- **Network Density**

  The network density as defined by de Nooy, *et al*. (2005) is the percentage of actual lines present in the network to the maximum possible number of arcs and it depends on the size of the network. The higher density indicates higher number of connections among the nodes, more interaction between the scientists, leading to a tighter structure and a more cohesive network.

- **Betweenness Centrality**

An actor's potential control of communication within the network can be indicated by betweenness centrality (Chung, and Hossain, 2009; Abbasi, and Altmann, 2011). It is defined as the ratio of the number of shortest paths (between all pairs of nodes) that pass through a given node divided by the total number of shortest paths. The highest betweenness centrality suggests the most central vertices. In other words, vertices (authors) with high betweenness centrality play critical role in the knowledge transmission between different nodes that are directly connected to the most central

[15] One-mode networks are networks where we study how all actors are tied to one another according to one relation, like friendship.
[16] The relationships in undirected networks represented by undirected ties (edges) because both individuals are equally involved in the relation.
[17] Excel2Pajek is a windows program developed in Delphi 7 by Jürgen Pfeffer, from FAS.research, Vienna to convert Excel datasets into **Pajek** format. It can be downloaded from: http://vlado.fmf.uni-lj.si/pub/networks/pajek/howto/excel2Pajek.htm

ones. Our concern in this measure comes from the suggestion by de Nooy, *et al*., (2005) that targeting the actors with highest betweenness-centrality is a good strategy for launching an innovation.

- **Degree Centrality**

Degree centrality is an indicator of an actor's communication activity (Chung, and Hossain, 2009; Abbasi, and Altmann, 2011). In a simple undirected network the degree of a vertex specifies the number of its neighbors. Likewise, the degree of each vertex, which represents a researcher, indicates how many collaborators he/she used to work with.

- **Weighted Degree Centrality**

The weight of the link wij between node i and node j indicates the strength of their collaboration tie, which reflects how many times they have repeated the collaboration. We calculated the weighted degree for each author by dividing the sum of their link weights (total number of co-authorships) by the total number of different co-authors. Scholars with a strong relationship (frequent co-authorship with the same partner) are considered as loyal ones (Abbasi, and Altmann, 2011).

- **Clustering Coefficient**

The clustering coefficient (CC) of a vertex (node) in a network graph quantifies how close its neighbors are to being a clique[18] (complete graph). In other words, it shows how related each scientist is to his/her neighbors, and the probability that they become a closed research group. Clustering coefficient is simply the number of edges between the neighbors, divided by the maximum possible for the type of network, k (k -1) or k (k-

[18] Based on the graph theory a clique in an undirected graph is a subset of its vertices such that every two vertices in the subset are connected by an edge.

1)= 2. It is worth mentioning that the clustering coefficient is decreasing over the years, with around 20% chance of two scientist collaborating if both have done so with a third scientists (Perc, 2010).

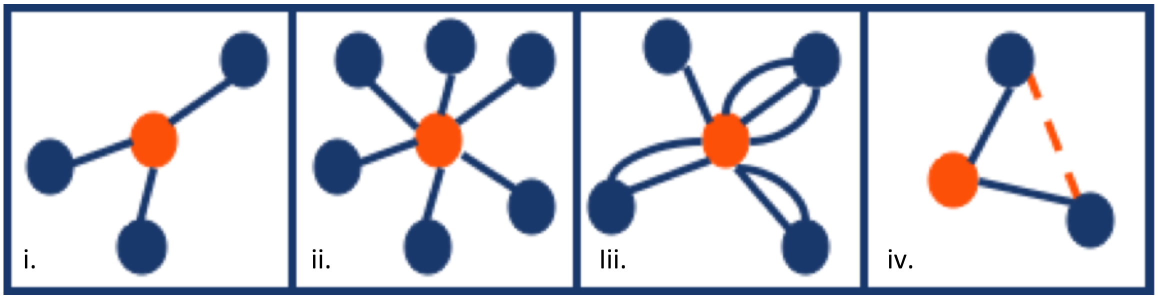

Figure 5: The network representation for different properties: (i) Betweenness centrality (ii) Degree centrality (iii) Weighted degree centrality (iv) Clustering coefficient

4.3 Data Analysis

As mentioned before, we have used the proper SQL quires to extract only the articles coauthored by Canadian scientists from the original database created by Moazami (2012). A total of 81,727 articles coauthored by 21,498 scientists have been extracted. For the purpose of analysis, we have build a new database with a complete record for each scientist containing information about his/her location, affiliation, research performance indicators and social network measures. (See the database dictionary in Appendix I).

4.3.1 Data Mining and Statistical Analysis

We will perform data mining techniques and statistical methods. On one hand data mining was performed through exploratory data analysis and extreme value analysis. On the other hand the statistical methods that were used were hypothesis testing and statistical distribution. The objective is to understand the behaviour of scientists in real world and to detect a pattern for each group of scientists in our database in terms of research performance and collaboration activities. All the techniques and methods used in this section of the study were performed using the RapidMiner[19] software and MS Excel was used for presentation purposes only.

4.3.2 Data Validation and Transformation

Data validation is not what we can call a data mining nor a statistical method but should be performed before any such activities is undertaken to ensure coherence and consistency. As an example, it was observed that the firm variable had seven categories; two of them were Lab and Laboratory. These two categories were merged into one, which is the Laboratory segment. The categories were reviewed down to five; Laboratory, Hospital, Industry, Research Institute and Academia. And the categories were transformed from text variables to categorical numeric, from 1 to 5, for modeling purposes.

We also used the Extreme Value Analysis (EVA) approach in order to deal with the extreme deviations from the measures of central tendencies. EVA seeks to assess, from a given ordered sample of a given random variable, the probability of events that are more extreme than any previously observed. In our study we decided to remove extreme observations from our databases before performing the statistical analysis. A good example is the following. As we explored the given variables we observed that 15 observations on the total number of citations were extremely high, over 2 billion. We

[19] **RapidMiner** is a software platform developed by the company of the same name that provides an integrated environment for machine learning, data mining, text mining, predictive analytics and business analytics. It is used for business and industrial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the data mining process including results visualization, validation and optimization. RapidMiner is a free open source program and is available online at : http://rapidminer.com/download-rapidminer/

thus dug deeper to find out that all these 15 observations had all the same values and way over the second next higher observation of about 26,500 citations. We decided to remove these 15 observations from our study.

4.4 Survey for Influences on Partnership Decision

4.4.1 Web-Based Survey As Research Methodology

As our network created by socially connected individuals and considering that human beings may have very personal motivations and reasons for choosing to work with others, a direct questionnaire for each individual scientist has been conducted for detecting the collaboration motivations. Shading some light on the intentions of the scientists through developing a questionnaire would help us understand the partners' selection mechanism (Alizadeh, 2011). As a complementary data collection approach we ran a survey sent to active researchers identified in our database as having scientific collaborations. The main objective was to elucidate the personal preferences to be considered while seeking potential collaborators for conducting a research project.

Web-based survey as a tool for online data collection is becoming an increasingly widespread research methodology. It has some advantages over paper-and-pencil surveys such as reduced time, lowered cost, ease of data entry, flexibility in format, and ability to capture additional information (Granello and Wheaton, 2004). An additional advantage is the ability to access individuals all over Canada and also those who used to reside in Canada when they published the articles so they are included in our database.

A potential limitation for this method that was shown by several studies is the significantly low response rate (Granello and Wheaton, 2004). However, some researchers discussed this issue in more details. For example, Crawford *et al.* (2001)

found that response rates increased when participants were told in their initial e-mail how much time the survey would take. Moreover, technical difficulties and formatting issues can lead to the loss the interest of answering the questionnaire and consequently to lower the response rate (Bosnjak and Tuten, 2001).

To avoid that, we tried to select the best qualitative question type, which is both convenient (easy and fast to complete) and can handle our objective as accurately as possible. For that purpose, we used the scaling measurement 'Likert[20]' to measure the scientists' attitude and its impact on their behavior. Among different scaling methods, the Likert-type question is more direct to answering the research questions meaningfully (Likert, 1932). It is the most commonly used approach to accurate scaling responses in survey research for measuring either positive or negative response to a statement.

4.4.2 Developing the Questionnaire

The questionnaire (Appendix II) included 18 factors that are related to the collaborative research activities. Using Likert approach, we have provided a range of responses for each question regarding the importance of this set of factors in influencing the scientist decision while seeking their collaborators. Five ordered responses levels were used including the extreme negative, extreme positive and neutral option. The options from weakest to strongest are: unimportant, slightly important, important, very important and critically important.

Since our survey is not public and specified personals that are identified in our database

[20] The scale is named after its inventor, psychologist Rensis Likert. Likert distinguished between a scale proper, which emerges from collective responses to a set of items (usually eight or more), and the format in which responses are scored along a range.

should answer it, we were expecting a very low response rate. In order to avoid that, we took the previously mentioned points into consideration while designing and sending out the survey. We have mentioned in the email that the survey is very short and should not take more than 5 minutes. In addition, we tested the page using different operating systems and web browsers to ensure that it is easily downloaded and maintains its formatting in all types of software and hardware environments. Besides, Concordia ENCS account has been used to send personalized email for each individual in our sample so they would not consider it a spam and trust that the purpose behind the survey is indeed a scientific research.

4.4.3 Population and Sample Determination

The population for our survey is the nanotechnology coauthors that are affiliated to Canadian institutions. As mentioned previously, our database consists of 21,498 scientists who have published at least one article in the field of nanotechnology during the study period. While determining the sample size we allow +/- 5% a margin of error at 90% confidence level. Based on these assumptions we required a minimum sample size of 271. In order to have sufficient responses, we have randomly selected 1,500 researchers from our database to be surveyed. The sample contains researchers from different provinces, distinctive firms and having different research performance. Participants were recruited on a voluntary basis through email and were asked not to forward it to others because participation in this survey is limited to specific researchers identified in our database. We have collected their contact information from the official websites for the organizations they are affiliated to. Afterwards, we have designed an online survey using Google forms, which has been sent to those scientists through emails.

4.5 Simulation Model Building

4.5.1 Agent-Based Modeling (ABM)

An agent-based model consists of a system of agents that are repetitively interacting and their dynamics can be explored using the power of computers (Epstein and Axtell, 1996; Axelrod, 1997 b; Bonabeau, 2002; Macal and North, 2010). Each object (agent) in ABM has several characteristics, which include a set of goals that are supposed to be accomplished, certain social behavior based on a set of social rules and interacting behavior with other agents (Weiss, 1999). Agents also are characterized by their learning capability and changing decision rules (Albino *et al*., 2003).

Agent-based models (ABM) are commonly used to represent individual actors (or groups) in a dynamic adaptive system (Garcia, 2005; Berryman and Angus, 2010). Moreover, they include representation of human behavior and are used to observe the collective effects of the agents' interaction among themselves and with their environment considering various factors (Goldstone and Janssen, 2005; Macal and North, 2010). Agent-based modeling is an effective tool in simulating the flows of scientific knowledge within collaborations (Scholz *et al*., 2010), effect of failure of partnership on the agent population, and how agent learning from partners and collaborators (Pyka *et al*., 2007).

The power of ABM in social science studies has been demonstrated by its ability to express a significant amount of data and knowledge about the behavior, motivations, and relationships of social agents. The flexibility of ABM is one of the most critical reasons for using this approach. The imperfect rationality, effects of learning and rules of interaction on the agents' ability to evolve, and social and institutional structure are some examples of the issues, which can be explored by applying the computational simulation in the social science (Bankes, 2002). ABM can facilitate various studies related to social networks such as modeling dynamically changing networks, capturing different types of agents and their behaviors (Berryman and Angus, 2010), exploring the diffusion of innovation and adoption dynamics (Bonabeau, 2002), and addressing the complexity of knowledge production processes in a manner not captured by more traditional research approaches (Pyka *et al*., 2010). ABM is an efficient research tool for studying the generation, distribution and influence of innovations within an industry (Schrempf, 2013).

Capturing the emergent phenomenon of a system is one of the essential benefits for using ABM, which has been mentioned in the literature. Some researchers defined the emergence as the exception rather than the rule (Bankes, 2002). Others like Bonabeau (2002) stated that the emergent phenomenon is the group's collective behavior, which is a result of simple individual rules and can be dramatically impacted by minor changes in those rules. Berryman and Angus (2010) and Goldstone and Janssen (2005), on the other hand, demonstrated these emergent phenomena as the events that occur when agents are interacting based on low-level rules, and sometimes they are really challenging to be observed without having a running model. Pyka *et al*. (2010) claims that it is possible to detect the collaboration patterns that generate the best emergence development using ABM, which also allows changing the settings of this process.

It is worth mentioning that the emergence of a system may result in a coevolving system where neighboring agents are directly influenced by an agent's change (Garcia, 2005). Agent-based models are useful in social systems when they are used to test a hypothesis to examine some possibilities, as well as to describe the system by "what-if" scenarios (Berryman and Angus, 2010; Barabási *et al.,* 2002). Nonlinear individual behavior such as when learning or adaptation occurs within the system is another case that ABM can be useful too (Bankes, 2002; Bonabeau, 2002; Garcia, 2005).

Other researchers also shed some light on the limitations and issues with using the agent-based modeling approach. Bonabeau (2002) for example mentioned three main problems with using this methodology. First, the agents' interactions are usually difficult to be modeled to match the reality since the agents are highly influenced by others around them. Human agents, in particular, have some soft or intangible factor (e.g. potentially irrational behavior, subjective choices, and complex psychology), which are difficult to be quantified, adjusted, and sometimes justified. Second, there is a difficulty of interpreting the quantitative outcomes of a simulation at the qualitative level. Lastly, ABM looks intensively at the constituent units of a system not only at the aggregate level, which can be a time consuming.

According to the literature, agent-based model is an effective approach to capture a very rich set of complex behaviors and interactions, and thus it is highly appropriate for modeling complex phenomena. Computer simulation can be used as experimental technique for hypothesis testing and scenario analysis, which can be performed complementary to and/or in combination with experiments in real-life, in the lab or on the Web. A properly designed model can deliver reliable results beyond the range of analytical tractability (Helbing, 2012). Several intentions to develop models have been discussed by Epstein (2008) including prediction, explanation, guiding data collection, revealing dynamical analogies, discovering new questions, illuminating core uncertainties, demonstrating tradeoffs, training practitioners, and decision support.

Agent-based simulations are powerful explanatory tool that can reflect interactions between different individuals. By modeling the relationships on the level of individuals in a rule-based way, agent-based simulations allow researchers to produce characteristic features of the system as emergent phenomena (Helbing, 2012). Indeed, ABM facilitates accomplishing complex objectives that cannot be done in traditional approaches, such as: the dynamic change in the network caused by its agents, learning and evolution of agents, and capturing a large range of different types of agents and their interaction behavior (Berryman and Angus, 2010). On account of these abilities, ABM has been used as the most appropriate approach to investigate the impact of several changes on the structure of Canadian nanotechnology knowledge-based network and its scientific production. For this purpose, we have developed a computer model simulating the collaborative behavior of scientists in our database using the Netlogo[21] package version 5.0.3 (Wilensky, 1999).

Netlogo is a commonly used simulation platform for agent based modeling by individual scientists or group of researchers. Real complex research projects can be found in Modeling Commons[22], Center of Connected Learning (CCL)[23] and SKIN

---

[21] **NetLogo** is a multi-agent programmable modeling environment. It is used by tens of thousands of students, teachers and researchers worldwide. It is authored by Uri Wilensky in 1999, and developed at the Center of Connected Learning (CCL) and can be downloaded free of charge at: http://ccl.northwestern.edu/netlogo/download.shtml

[22] The Modeling Commons is for sharing and discussing agent-based models written in Netlogo. It has more than 1,000 models, contributed by individual researchers from around the world.

[23] Northwestern's Center for Connected Learning and Computer-Based Modeling and the CCL research group headed by Prof. Uri Wilensky in 1995 at Tufts University and relocated to Northwestern University in 2000.

group websites. The acceptance of NetLogo in the research and education communities is wide and growing, which is evidenced by the tens of thousands copies of the software which have been downloaded. The NetLogo discussion group on yahoo has over 1,500 members and averages over 100 posts per month.

Over a number of modeling platforms such as Swarm, Mason, and Repast, we have chosen Netlogo for its power as a modeling environment, which consists of a programming language (derived from the earlier Logo language) and a set of libraries, as well as a programming environment. NetLogo is a standalone application written in Java and it is mostly suitable for modeling complex systems evolving over time with thousands of interacting agents. It can also exchange data with other applications to let you read or write any kind of text files.

Moreover, it provides the ability to include a wide range of chosen parameters to capture the complex phenomena. Another useful feature of NetLogo is that it provides a graphical tool for quickly constructing interfaces for running agent-based models. The contents of the graphics window, or of the model's whole interface, can be saved as an image and can be exported to Excel files for further statistical analysis and clearer graphical representation.

Beside the technical capabilities, the existence of the extensive documentation and tutorials as well as the availability of large collection of open source codes in Netlogo has supported our decision to use it. NetLogo's Models Library, for example, has more than 140 pre-built simulations that can be explored and modified (Tisue, and Wilensky, 2004).

4.5.2 Model Description and Assumptions

The knowledge-based innovation network can be thought of as a complex system with many interacting entities under certain environmental factors. Such complex systems consist of heterogeneous, adaptive and localized agents who act autonomously by assessing their state and making decisions to collaborate with others. These decisions are made based on a predefined set of rules and might include a certain degree of randomness. As stated by Dawid 2006 “The modeling of the dynamic interaction between individuals who might be heterogeneous in several dimensions and whose decisions are determined by evolving decision rules can be readily realized in ABM models”.

Our model simulates the knowledge creation and exchange interactions among a set of agents that act in a complex and changing environment given some rules and initial conditions. Its agents are the scientists identified in our database as the ones who have published in nanotechnology at least once with a Canadian affiliation during our study period. These scientists try to interact with others who are also seeking partners to conduct collaborative research projects and publish new articles. In the next sections, the elements and processes of our model are described in further details.

- **The agents and links**

Around 14,000 Canadian scientists in our database, who are also the nodes of the network, act as the individual agents of our model and are characterized by a set of parameters reflecting their research performance, scientific collaboration activities and network properties as in 2012. The following table presents all these variables and their description.

Category	Parameter	Description
Identification and Status	Node-ID	The author identification number as in SCOPUS
	Firm	The category of author's current affiliation as in 2012
	Star?	True when this author is a star scientists
	Gatekeeper?	True when this author is a gatekeeper
	Popular?	True when this author is a popular scientist
	Loyal?	True when this author is a loyal scientists
	Embedded?	True when this author is well connected to others in the cluster
Research Performance	Nano-articles	Number of the author's publications which contain the specialized keywords in nanotechnology
	All-Articles	Number of all articles that the author has in SCOPUS
	Citation-Count	Total number citations this authors' articles received
	H-index	The H index considering SCOPUS articles published after 1995
Collaboration	Max-partners	The maximum number of potential partners the author may search for
	Previous-partners	Agent-set of authors with whom the author has previously partnered
Network Properties	Betweenness	Betweenness centrality of this node in the network
	Degree	Degree centrality of this node in the network
	wDegree	Weighted degree centrality of this node in the network
	CC	Clustering coefficient of this node in the network

Table 3: List of model parameters owned by its agents (authors)

The initial values for these parameters will be loaded into the model through reading text files created based on proper SQL quires from our database. The first text file contains identification, research performance and network properties data for each author.

The Node-ID (Author ID) and the research performance indicators, as extracted from SCOPUS, will be given for each node to represent an author. On the other hand, the network properties for each node will be set up according to the previously performed social network analysis based on the network structure by the end of 2012.

The second text file contains information about the collaborative activities history. It consists of the co-authorship relationships between each two scientists that will be represented as links in our model. Each link has a weight reflecting the strength of their collaboration relationships based on how many times that have coauthored an article together. All scientists who have prior collaboration with an author will be stored as his/her previous-partners agent set. Considering that this is a two-way relationship, the pair of scientists at both ends of each link will be added to each other's previous-partners agent set. This group of scientists, who consist the previous-partners agent set, will be referred to while seeking partners for new collaboration as it will be discussed later in this chapter (partnership section).

- **The Environment**

Within the model there are two groups of global variables for setting the environment. The values for the first group of variables will be given using the sliders on the interface and they determine the percentage of scientists in each group to the whole population. The default value for each group is 5%, while we will decrease and increase this percentage in different scenarios for analyzing the effect of this change on the structure and efficiency of the network. The initially given value for each of the status parameters (i.e. *Star?, Gatekeeper?,..* ,etc.) is false, and will be changed to true for a ratio of the scientists with the highest values for the associated parameters.

The interface switches set the second group of variables to represent the existence of each group in the world. All switches are set to ON by default, which means scientists belonging to all groups exist unless other settings are specified. When a switch gives a false value, the nodes representing the scientists in the associated group will die (the node will be removed completely from the world along with the collaborative links the author entertained). The purpose of using this setting is to examine the role of each group of scientists by investigating the impact of their absence on both network structure and productivity. While the model is running, these two settings (the ratio of each group and whether they exist or not) will be implemented at each model's iteration considering the updated agents' variables from the previous iteration.

- **Partnerships**

An agent in the model may consider partnerships and start seeking potential partners to collaborate with in order to complement their knowledge and consequently publish a new article. For each iteration (time unit), a random number of nodes will be acting as starters who will initiate the partnership process by searching for candidates to collaborate with. In experimenting with the model, starters will follow different strategies for seeking their partners and another starter also can select them. In other words, an agent can be involved in more than one collaboration activity at the same time with a maximum number of partners for each involvement.

A potential partner who has a satisfactory prior collaboration experience with an author will most likely attract him/her for a new one. This is reflected in the model by the *repeat collaboration* function: to find a partner, the author will search among previous partners agent set and assign some as candidates. The number of candidates should not exceed the maximum allowable number of partners (more details about the number of partners will be discussed later in the model verification part).

The most centralized nodes in the network will be also attractive to be selected as candidates for new collaborations. That is, star scientists and gatekeepers will be most frequently selected over others to act as potential partners. The group of starters who will follow this strategy in the model will search among those agents who have given true values for their *Star?* and *Gatekeeper?* variables during setting the environment as discussed earlier. The rest of starters will be open for new collaboration with any available agent.

After finding the candidates, the partnership relationship will be established, where for some of them it will be based on the preferable number of partners according to past collaboration. If this is the first time for a pair of scientists to collaborate a new link will be created between them and a value of 1 will be given to its strength. Alternatively, if the collaboration tie between them already exists, its strength will be incremented by 1. That is, 0.5 for each side of the relationship to avoid the redundancy. We are assuming that each collaboration activity is resulting in a new publication coauthored by the involved scientists. Thus the variable (*Nano articles*) for each of these agents will be also increased by 1. Besides, the actual partners will be added to previous partners agent set for a future collaboration that might occur in the next iterations.

- **The Networks**

Only agents that have participated in any collaboration activity during this step (iteration) will be given an age value equal to the step number x. These agents will form the new network whose structure and productivity will be examined. For all nodes with (age = x) we will recalculate the values of variables related to network measurements. The Netlogo NW[24] extension for network analysis have been integrated with our model to reanalyze the network in each iteration based on the new collaboration activities. The degree centrality, betweenness centrality and clustering coefficient for each node in the

[24] NW is an extended library that can be integrated with models developed in Netlogo to perform the social network analysis. More information and the downloadable files are available at: https://github.com/NetLogo/NW-Extension

new network will be updated as values for the associated variables. After updating the values the structure measurements for the whole network will be calculated by averaging the values of individual participants. Before moving to the next iteration randomly selected agents who were a part of this network will be completely removed from the network. This represents the behavior in the real world where some scientists publish only once and quit the network after.

According to the changes in the performance and centrality of scientists involved in lately formed network they might have different status and act as new or different member of the identified groups. That will be verified by implementing the set up world functions at the beginning of each iteration. That will find the agents with the highest values for the associated variables and change their status parameters to true and remove the scientists in specific group if any of the switches is set to OFF.

The flowchart below describes the sequence of the process in the developed model.

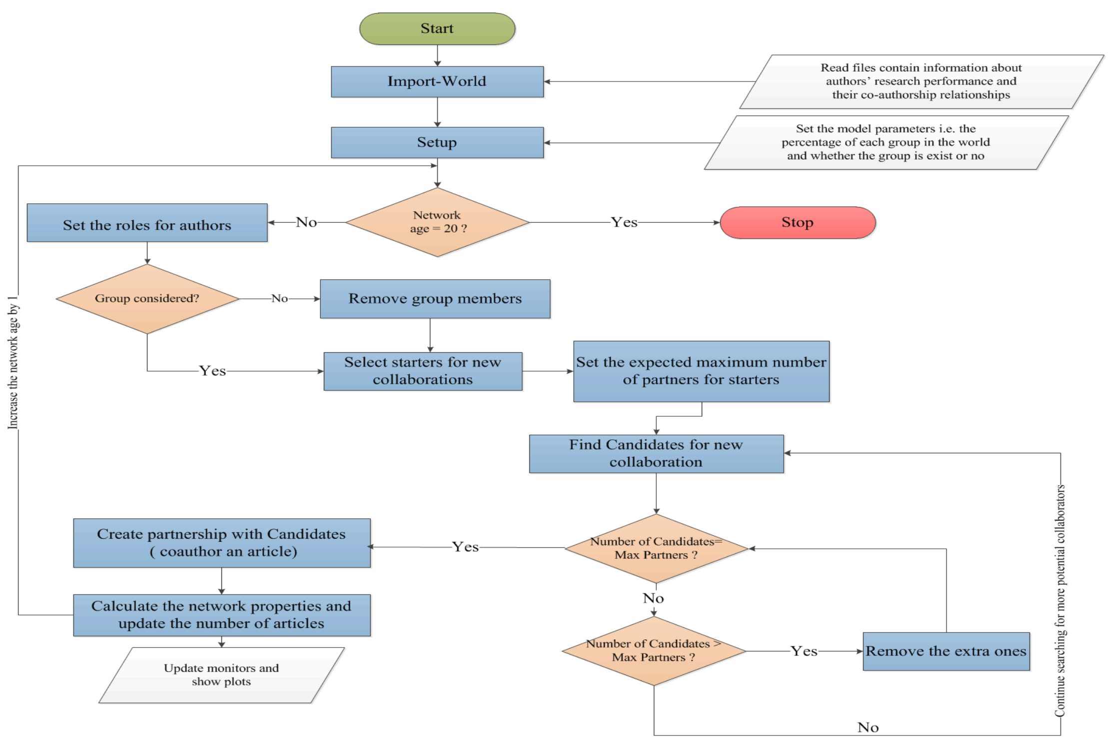

Figure 6: Flowchart of the developed simulation model

While building the model we have considered several conceptual assumptions. First of all, the collaboration strategy and individuals behavior in the model reflect the result of our extensive analysis for the historical data collected from SCOPUS and the questionnaire. For seeking the potential partners, over 50% of scientists prefer to renew their collaboration relationship with the researchers with whom they previously worked. This is implemented in the model by asking half of the nodes to search among their previous partners agent set when they are ready for a new collaboration. Other attractiveness indicators, based on our survey, are the reputation of the partner in the field, resource accessibility and having a common research interest. This selection mechanism for establishing new partnership has been implemented by referring to the detailed information about each author. The most attractive scientists for a new partnership are gatekeepers, star scientists and those who have H-index greater than or equal 17 (which is the average for the best performing scientists according to the data mining analysis).

Moreover, the maximum number of potential partners has been determined referring to the degree probability analysis of the database. Based on the probability density function we have found that the highest likelihood is to have no more than 10 partners. Accordingly, we have assigned 10 as the maximum allowable candidates that an author will search for, while each will have an actual partnership with the preferable number the model learned from the collaboration history.

Finally, we assumed that each starter would seek to establish a new partnership, which will result in at least one new publication. Giving the change rate in the publications volume over the study period (see table 1 and figure 4), where the considerable increase after 1996 is detectable; the model has been programmed in a way to represent this evolving trend. We asked the number of starters to increase by a random percentage between 1.34 and 2.54 every year, thus the outcome will be increased by a ratio corresponding to reality.

4.5.3 Model Verification and Validation (V&V)

In the case of agent-based modeling, both the behaviors of agents and their software implementations must be verified. The importance of validation and verification is to ensure that the program code faithfully reflects the behavior of the conceptual model[25] as well as to determine whether the model and its results are valid for a specific use or purpose (Sargent, 2013). Quantifying the predictive accuracy of the model is an expected outcome of the model V&V process (Thacker *et al*., 2004). In the present section we are intending to ensure that the model and simulation result correspond to the real world as accurately as possible.

- **Computerized Model Verification**

Using special-purpose or general-purpose simulation languages will generally reduce the probability of coding errors comparing to when a general-purpose higher-level programming language is used (Sargent, 2012). As we have used Netlogo, which is a specialized package for developing ABM models, our model verification is primarily concerned with ensuring an error-free model that has been programmed correctly in the simulation language. Moreover, we have used some verification techniques to ascertain code correctness and robustness such as structured walkthroughs and traces. White box

[25] The conceptual model is developed by modeling the system for the objectives of the simulation study using the understanding of the system contained in the system theories.

testing at unit and integration levels have been performed to ensure code optimization by revealing hidden errors and minimize the amount of memory occupied.

- **Conceptual Model Validation (Internal Validity)**

Validation is the process of determining the degree to which a model accurately corresponds to the real world from the perspective of the intended uses of the model (Thacker *et al.*, 2004). The conceptual model theory has been built with the highest possible level of accuracy and correspondence to the real world behavior.

We have considered making a large number of simulation runs, so the confidence interval will narrow to the rate $1/\sqrt{n}$ in addition to the t-statistic that is calculated becoming smaller because the tails of the t-distribution get less fat as the degrees of freedom increase. To estimate the number of replications we need to perform in order to achieve a desired precision, we first measured the precision using the absolute error (Currie and Cheng, 2013). We have found the expected number of replications that are needed for the estimator to be within $100\varepsilon\%$ of the true value with confidence level 95% is at minimum 10.

For measuring the system's performance, it is essential that more than one run of the simulation model be used to generate the results (Currie and Cheng, 2013). The amount of variability in the model has been determined by running several replications for each experiment where the smallest amount of variability indicates the best performance (Sargent, 2013). We have carried out several runs or replications and used the means of the output performance indicators as our prediction for their values. We were expecting a slightly different data produced in each replication due to the sort of randomness we included for selecting the number of starters. However, the variance between the different runs results should not be high. The following figures show examples of output data over for twenty years for 10 independent replications. We can see from the figures that the trends are similar, which confirms that the model is working homogeneously.

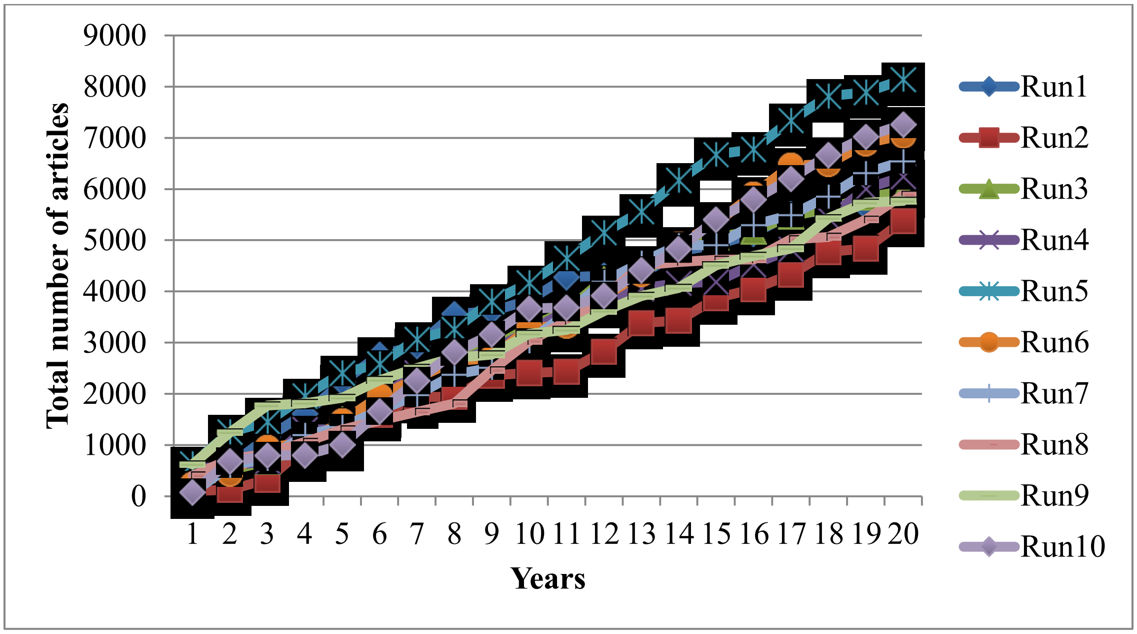

Figure 7: Internal validity. Data trend for number of published articles in the network

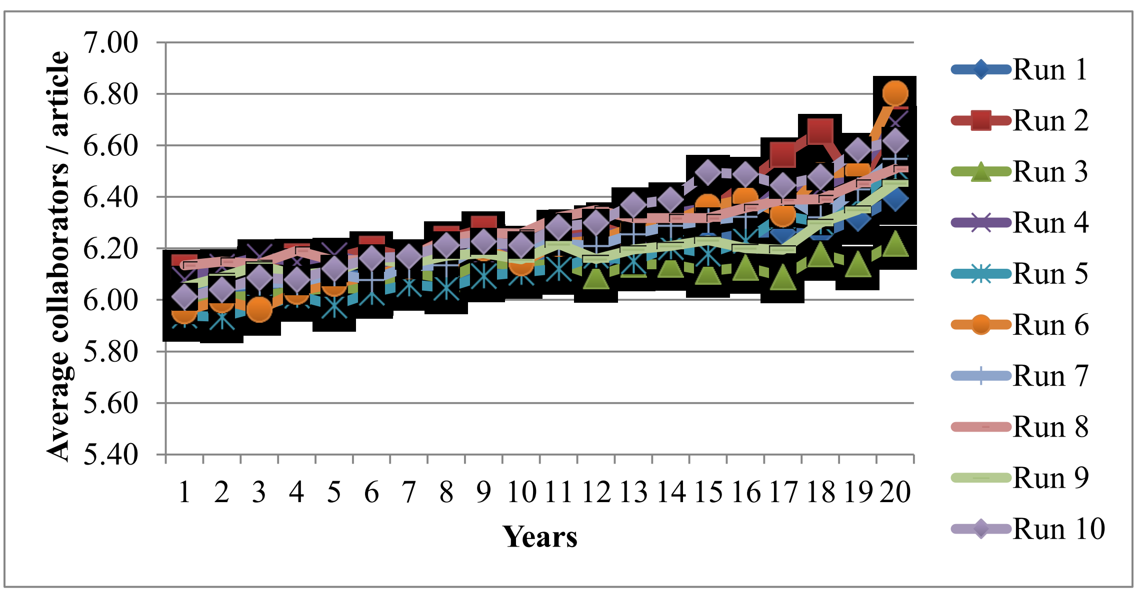

Figure 8: Internal validity. Data trend for number of coauthored per article in the network

- **Operational Model Validation (External Validity)**

Sargent (2012) discussed several techniques and tests commonly used in the literature for operational validation. We have used a combination of these techniques for validating the sub-models and the overall model as follows:

- **Animation Validation:** The objective of this technique is to ensure that the external appearance of the model matches its concepts. A graphical representation for the model's operational behavior is displayed as the model moves through time. Our model visualizes the network structure by distributing the agents in the world first and then create the links between them based on the data read form the text file. Besides, the interface displays of a number of monitors and plots that represents the updates in the variables of interest values while the model is running. The figure below shows the model's interface to give an impression of its performance in the scenario where standard parameter settings are used.

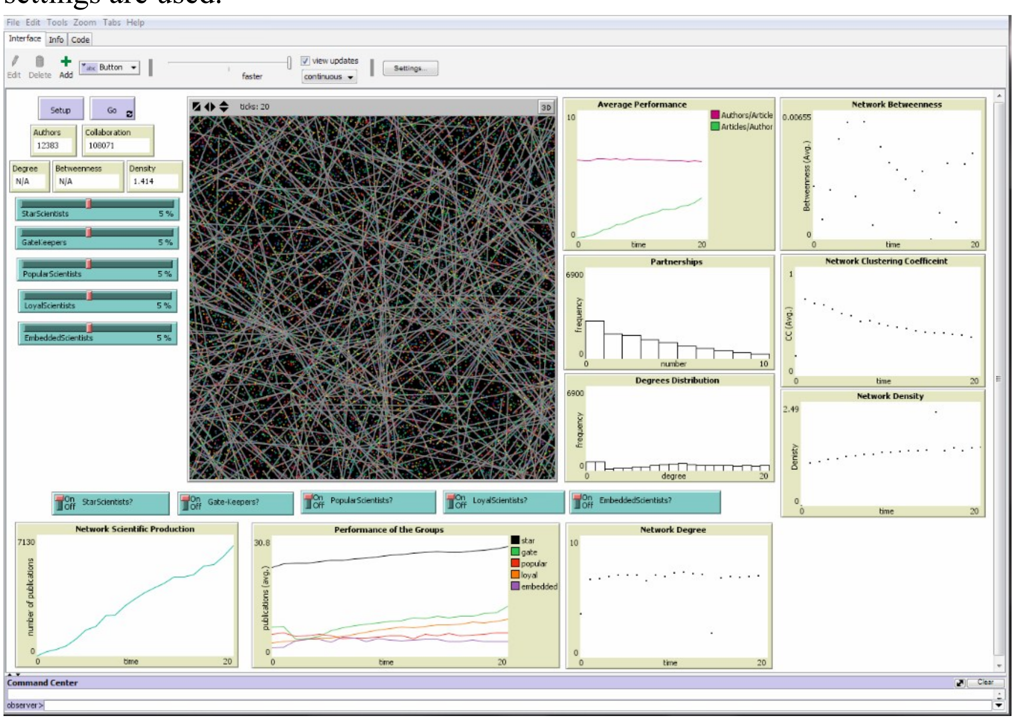

Figure 9: The model interface

To start working with the model the (setup) button should be pressed which will execute the initial commands. This includes clearing all the monitors and plots, then start reading the first text file that contains the authors' information. While reading the file, the agents will be created and their information will be stored in the associated variables. Meanwhile, the authors' monitor will be updated to mirror the authors' count. The main graphic window (Figure 9) shows all the Canadian authors in the text file (represented by the small default colored shapes). Their position in the display window is not significant: a layout algorithm is used to move the author icons to positions where they can best be seen. When the model reaches the end of the first file, the links information will be imported from the second file and the connections between the agents will be created based on their collaboration history to indicate their partnership. The count of links created will be appear as an update for collaboration monitor where it is strength will be stored as a value for the proper variable.

The graphs and plots surrounding the display window monitor various aggregate aspects of the system. Regarding the system performance, the first graph at the left bottom corner ('Network Scientific Production') shows the total number of publications produced by all authors each year. The next one ('Performance of the Groups') shows the productivity of each group of scientists by averaging the number of articles in nanotechnology that scientists in each group published. At the top right, close to the display window, the ('Average Performance') plot demonstrates the change in the average number of authors in each collaboration activity as well as the average number of articles in which each author participates at the same time. The two graphs below ('Partnerships') and ('Degrees Distribution') exhibit a histogram for the frequency of articles coauthored by the same number of scientists and the frequency of authors having the same number of collaborators respectively.

Concerning the network structure, at the very top right corner the ('Network Betweenness') displays the average betweenness centrality of authors involved in collaboration activities in each model iteration. Similarly, the graph below ("Network Clustering Coefficient') and ('Network Degree') correspondingly plots the average clustering coefficient and the average degree centrality. The network density is calculated by finding the portion of the actual connections according to all possible ones considering the number of network nodes and is presented in the ('Network Density') graph. In addition, the three monitors on the left of the screen display the correlated values and update them after each iteration. Based on our previous knowledge in the real world system, we can conclude that the values in the monitors and plot trends indicate a reasonable behavior for the model.

- **Comparison Validation:** Various results of the simulation model are compared to known results of our analytic model. Since we could analyze only the simple case, the comparison will be for the result of the scenario with the default settings. A graphical representation is used to compare the data of the simulation model and system output variables to determine the accuracy of the model's output behavior.

  The figure below shows a histogram comparing the average number of articles by each group of scientists in both the real data analysis and the simulated system. The identical results indicate a satisfactory range of accuracy for the simulation model behavior.

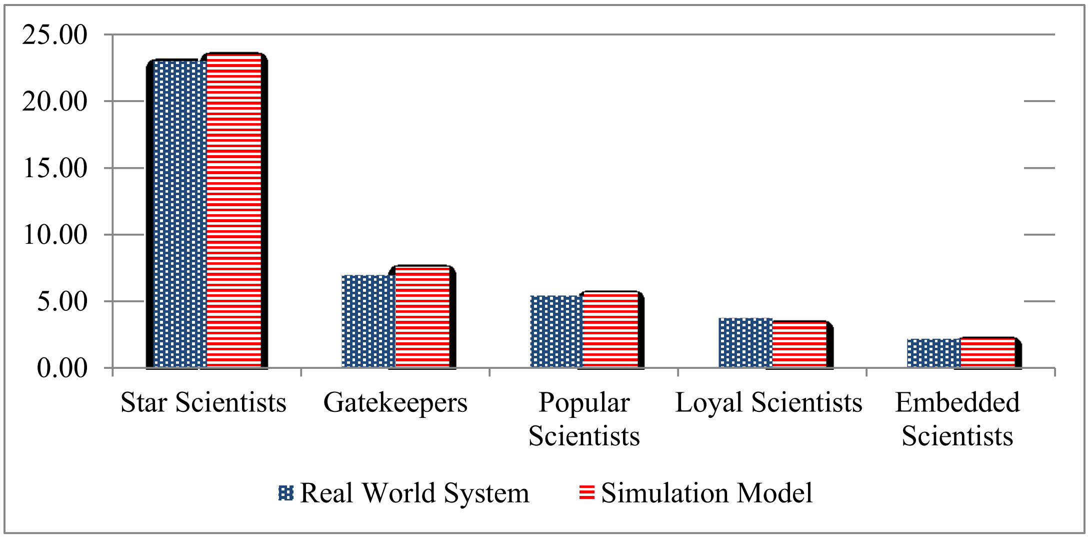

Figure 10:The average performance per group in analytical and simulated models

- **Degenerate tests:** We have tested the degeneracy of the model's behavior by the appropriate selection of values of the input and internal parameters. For example, the gatekeepers who have been excluded from the network. It is expected that in this scenario, the knowledge will remain inside the cluster and the research groups will be more closed and consequently the performance of embedded scientists will be improved. While applying this setting, as it is expected, the average number of articles published by embedded scientists has increased over time. The Figure below depicts the performance of embedded scientists in both scenarios (with the existence and absence of gatekeepers) over twenty years. The result from this example verifies a considerably acceptable level of our model's behavior degeneracy.

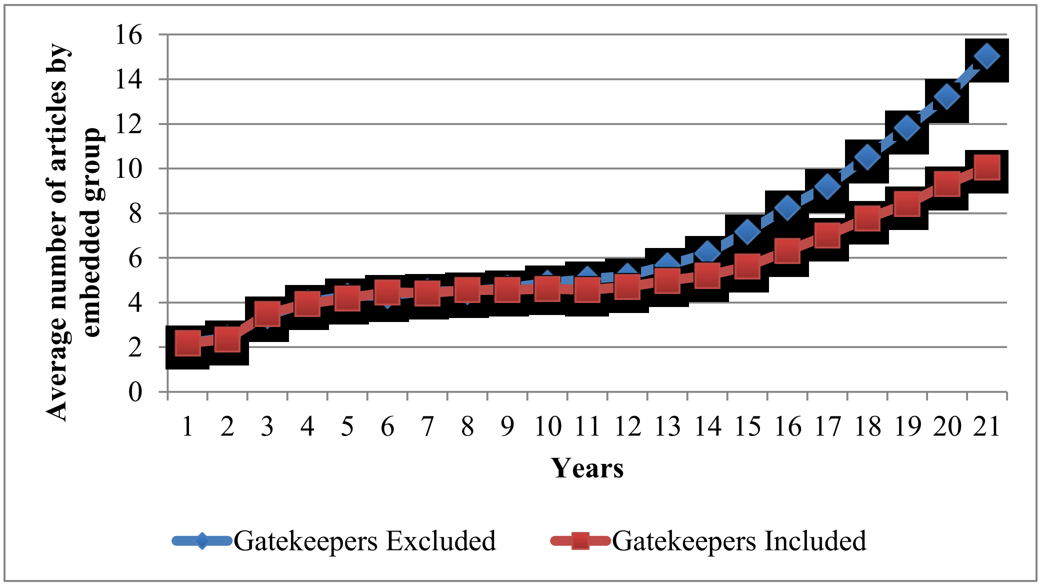

Figure 11: Degenerate test. The improvement in the productivity of embedded scientists in case of gatekeepers' absence

4.6 Experimental Scenarios

Indeed, the introduced groups of scientists in this thesis are parts of the knowledge-based networks, that is, they appear and grow in the networks naturally. Accordingly, the hypotheses regarding their number as well as their absence from the network can by justified only through simulated scenarios and not by real evidences. To examine the impact of the absence of each group on the production and the structure of the network, a substitution for the real world would be required.

The parameter variability analysis is implemented by carrying out several experiments to examine the effect of changing the values of the input and internal parameters of the model upon the model's behavior or output. Various scenarios are simulated to study the role of each group of scientists first by removing them completely from the network and then by increasing and decreasing their ratio to the population.

Using BehaviorSpace[26] we have run the model many times, systematically varying the model's settings "parameter sweeping" and recording the results of each model run. Beside the basic scenario where each group is present as 5% of the population, the experimental scenarios used two values for the switches (true and false) reflecting the existence and absence of each group respectively. The objective of these scenarios is to examine the role of scientists representing each group by removing them completely from the network. In other words, in each scenario we have removed the nodes that act as specific group along with their links (i.e. their collaboration ties will be removed also, but their partners will remain in the network open for new partnerships).

The other set of experiments used four different values for each slider reflecting the increase and decrease of the group's ratio to the population (2 scenarios each). Since we used 5% as default setting, we used 1%, 3%, 7% and 9% as testing values.

In each scenario, 20 iterations of the model are run, which represents the change in the values of interest over 20 years. We have used 10 replications of each experiment, and the results are then averaged for these ten runs of the model. We have examined the change of one value only while the rest of the settings remain the same. For comparing and evaluating the scenarios we are mainly concerned about the performance and the structure of the network.

As for the performance, the number of publications for the whole network and the average of the articles published by each group are used as indicator of the productivity. On the other hand, we have examined the structure of the network as it plays the key role in the diffusion of knowledge and production of innovation. The network structure

[26] BehaviorSpace is a software tool integrated with NetLogo that allows you to perform experiments with models.

properties have been calculated by averaging the values of the corresponding variables for all nodes that the network consists of. Degree centrality, betweenness centrality, clustering coefficient and network density have been calculated and compared in the different scenarios to evaluate the impact of the changed setting.

5.0 Analysis and Results

5.1 Network Visualization and Mathematical Analysis

The structural features of the network have been summarized as a single number in the general network structure report shown below:

Number of vertices (n): 21,498	
Total number of lines (Edges)	65,535
Number of multiple lines	29,568
Density [no loops allowed]	0.00029947
Average Degree	6.26499689
With summed lines Average Degree	4.00000000
Network Betweenness Centralization	0.05296156
Network Clustering Coefficient (Transitivity)	0.04994895

Table 4: The complete network analysis (Macro level analysis)

The analysis shows the number of all connections (edges) between the authors, where there are 29,568 multiple lines representing the repeated collaboration. The density of the network is 0.0003, which means that only 0.029% of all possible edges are present. The low density is actually expected in a network of such size, since the density is inversely related to network size. That is the larger the social network, the lower the density because the number of ties which each person can maintain is restricted comparing to the number of possible lines which increases rapidly with the number of vertices (De Nooy *et al.*, 2005). The average degree centrality of the network is 6.26, which means that each vertex is involved in 6 ties on average. A higher degree of vertices yields a denser network, because vertices entertain more ties. Average degree is a better measure of network overall consistency than density because it does not depend on network size, so it can be compared between networks of different sizes (De Nooy *et al*., 2005). The average degree with summed lines is 4, which indicates the average number of vertices that are connecting to this vertex: its neighbors. In other words, each author is adjacent to an average of 4 collaborators. Betweenness centralization, which is defined as the variation in the betweenness centrality[27] of vertices divided by the maximum variation in betweenness centrality scores possible in a network of the same size, is 0.0530. The network clustering coefficient of the network is 0.0499, which indicates the degree to which nodes in the graph tend to cluster together. In other words, it shows that there is a 49% likelihood that nodes tend to create tightly knit groups characterized by a relatively high density of ties; this tends to be greater than the average probability of a tie randomly established between two nodes.

Afterwards, we have calculated some common network measures for each vertex (author) in the network. These measurements include degree centrality, betweenness centrality, and clustering coefficient.

The minimum degree centrality for the vertices in our database is 1 that means there is no isolated vertex that has no neighbors. Based on the degree distribution, we used probability density function (pdf), or density of a continuous random variable, to describe the relative likelihood for random variable to take on a given value. The figure below shows the probability of the degree centrality falling within a particular range of

[27] Betweenness Centrality of a node measures the shortest paths between all the pairs of vertices present in the network which go through each vertex

values according to the original dataset used in the network analysis.

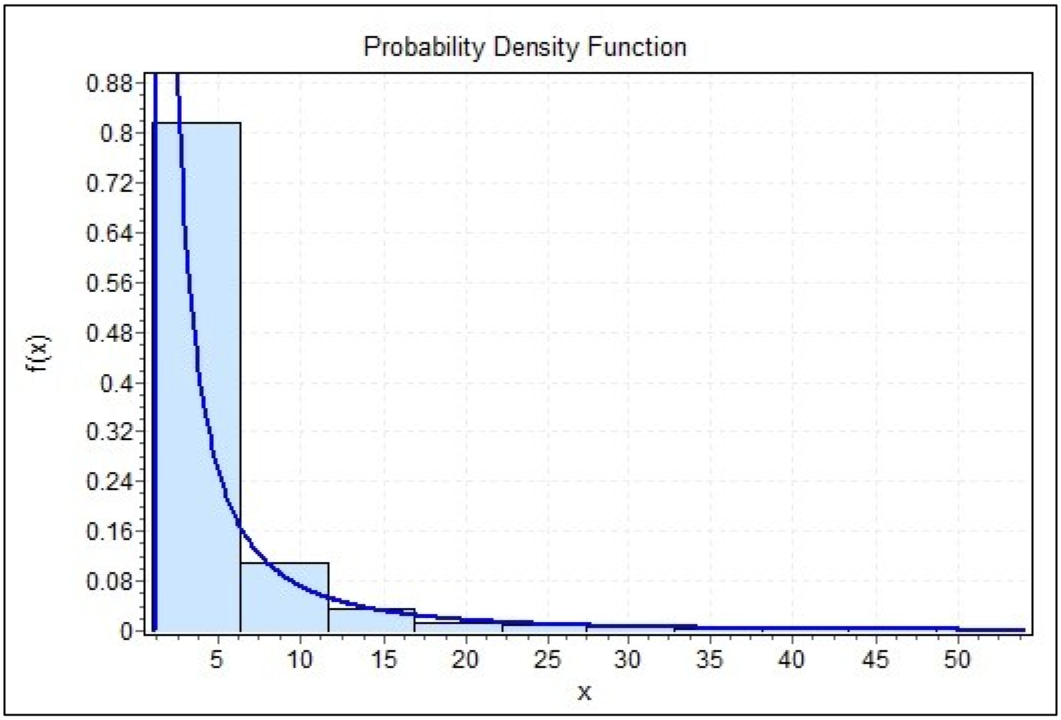

Figure 12: The probability density function of the network degree distribution

5.2 Characterize and Categorize the Scientists

In this section we intend to categorize the researchers in our database into groups to study the research performance and collaboration behavior for those who have common characteristics. We have proposed the social network measures to be used for the grouping purpose. In order to answer the question whether the position in the network and research performance of researchers are correlated we have statistically examined the relationship between the social network measures and the research performance indicators by calculating the correlation coefficient between these variables as following:

	Nano Articles	Citation Count	H-index	Betw-eeness	Degree	W Degree	CC
Nano Articles	1						
Citation Count	0.242	1					
H-index	0.332	0.831	1				
Betweenness Centrality	**0.108**	0.140	0.160	1			
Degree Centrality	**0.102**	0.083	0.110	**0.588**	1		
Weighted Degree	**0.045**	0.050	0.100	0.109	0.107	1	
Clustering Coefficient	**-0.040**	-0.048	-0.061	-0.009	0.292	0.018	1

Table 5: The correlation matrix between the performance indicators and network properties

The above table shows that the centrality measurements (betweenness, degree, and weighted degree) are positively correlated to the research performance, while the clustering coefficient is negatively correlated to the same. This result was expected as the higher clustering coefficient indicates more cliquishness and consequently slower knowledge flows and creation. The results also suggested a strong positive relationship between the degree centrality and betweenness centrality 0.588, which means that scholars who are centralized in the network have high number of connections and vise versa.

Based on the scientists' research performance information and the social network analysis measurements, we have identified five separate profiles of the researchers. The following sections introduce thee five groups of scientists and present the top ten scholars in each group in our database.

- **The Star Scientists**

The term "star scientist" has been used by (Zucker and Darby, 1996) to qualified researchers who improve research productivity by their excessive experience in research and innovative activities. In other words, star scientists are researchers with high impact on innovation and knowledge development reflected by their considerably higher productivity comparing to their colleagues and competitors.

Because of their knowledge that contributes significantly to the success of firms, star scientists are important in the process of technology transfer (Zucker, and Darby, 2005). Considering that star scientists are the scientists with greater number of publications, those researchers actually act as knowledge circulation improver and also as generators for new knowledge (Schiffauerova and Beaudry, 2008). In this thesis, we have assumed the 5% of the scientists in our database who have published the highest volume of articles are most productive ones and thus they are the star scientists in the Canadian nanotechnology industry. The star scientists coauthor around 40% of the total number of publications in our database.

Author	Nano-Articles	Citation Count	H-index	Location
Huichun Liu	**236**	2525	27	Ottawa
Zbig R. Wasilewski	**186**	3900	34	Ottawa
Randy D. Gascoyne	**171**	3712	34	Vancouver
M. Buchanan	**143**	1680	22	Ottawa
Theodore Cameron	**125**	2792	28	Halifax
Brian H. Robinson	**123**	3172	35	Toronto
Tomaiai Hudlicka	**119**	1307	22	St. Catharines
Kam Chiu Tam	**116**	1723	18	Waterloo
Brian D. Sykes	**110**	10962	43	Edminton
Yong Zhang	**101**	5158	44	London

Table 6: The top ten productive researchers with highest number of publications in nanotechnology

- **The Gatekeepers**

Targeting the actors with highest betweenness-centrality is an effective approach for launching an innovation (De Nooy *et al*., 2005). As mentioned before those scientists are the most centralized individuals who are formally responsible for providing the channels and link separate sources of knowledge. They are defined as brokers, and they are responsible for controlling the communication between scientists, who do not have either approach to or trust in each other (Marsden, 1982). Gatekeepers as suggested by Gould and Fernandez (1989), are the influential who are responsible for the knowledge transfer, and they are also valuable for merging different existing ideas that are held by their directly connected ones.

Gatekeepers symbolize those individuals who are bridging the information flows between two or more geographically separate clusters by making connections between them. Usually only maximum of one fifth of the innovators in the networks are accountable for transmittance of external fresh knowledge to a cluster (Schiffauerova and Beaudry, 2008). Accordingly, we have considered that those top 5% of all scientists in our network with the highest betweenness centrality to represent the Gatekeepers.

Author	Nano-Articles	Citation Count	H-index	Location	Betweenness Centrality
D. Jed Harrison	1	1117	18	Victoria	**0.032885**
Harry E. Ruda	5	752	11	Toronto	**0.023091**
E.H. Sargent	1	17134	59	Toronto	**0.017138**
Hicham Fenniri	7	408	12	Edmonton	**0.015787**
Jigang G. Zhou	6	150	9	Saskatoon	**0.014527**
Sandra Marcus	1	380	9	Edmonton	**0.014408**
Vincent Aimez	2	2348	30	Sherbrooke	**0.014226**
Hanan Anis	3	577	14	Ottawa	**0.014187**
Gregor Lawson	4	19	2	Hamilton	**0.014064**
Adam Hitchcock	2	3196	19	Hamilton	**0.013341**

Table 7: The top ten researchers with highest betweenness centrality

- **The Popular Scientists**

Alongside the previous research activities and background, the more links a scientist has to outside sources of knowledge, the more amount of fresh and new knowledge it can access and bring to his colleagues for further collaborative activities. Consequently, both the number of links each scientist has and their research positively affect the productivity rate of innovativeness in a firm (Henderson and Cockburn, 1996).

The more connections a scientist has indicate his/her higher connectivity in the network. Those researchers who are connected to a greater number of collaborators are critically important for sharing the knowledge, which lead to better scientific performance. As we used the degree centrality of each vertex (author) as an indication for their number of connections, we have considered the top 5% of scientists with the highest degree centrality as popular scientists who are sought-after collaborators and probably also very well known in the field of nanotechnology.

Author	Nano-Articles	Citation Count	H-index	Location	Degree Centrality
Larissa Levina	4	606	11	Toronto	**58**
Richard Soluk	10	1136	12	Edmonton	**54**
R. McPherson	1	5511	39	Victoria	**54**
M. Wang	1	59	1	Regina	**54**
Dan Tzur	1	300	7	Edmonton	**46**
Ruying Li	3	1579	19	London	**45**
D. Jed Harrison	1	1117	18	Victoria	**44**
Dean Cheng	1	72	6	Edmonton	**44**
Kenvin Jeroncic	1	66	2	Edmonton	**44**
Summit Sawhney	1	99	3	Edmonton	**44**

Table 8: The top ten researchers with highest degree centrality

- **The Loyal Scientists**

Several studies suggests that partner's selection process is time consuming and thus scientists prefer to remain loyal to their previous partners, and these studies also examined the impact of their loyalty on the overall productivity (Van Segbroeck *et al*., 2009). The general result concluded that maintaining previous partnership relationships has an impact on the network performance. We aim in this thesis to investigate the impact of scientists' loyalty on both network productivity and structure. As for the previously defined groups, we have considered the top 5% with the highest weighted degree as the most loyal scientists among our database.

Author	Nano-Articles	Citation Count	H-index	Location	Weighted Degree
Malcolm Xing	1	18	3	Winnipeg	**31**
Fartash Vasefi	1	693	14	London	**27**
J. F. Cochran	10	1971	23	Burnaby	**26**
Gabriel Devenyi	1	31	3	Hamilton	**26**
Wen Zhong	1	53	5	Winnipeg	**20**
Mohamadreza Najiminaini	5	514	15	Burnaby	**18**
J. Koropatnick	1	411	11	London	**17.8**
Reggie Hamdy	4	465	15	Montreal	**17.5**
Eric Martineau	1	45	3	Ottawa	**17**
Chris Payette	5	367	8	Montreal	**16**

Table 9: The top ten researchers with highest weighted degree

- **The Embedded Scientists**

We have defined the embedded scientists as those who are much willing to collaborate with the neighbors of their neighbors, which can be measured by their high clustering coefficient. The higher clustering coefficient a vertex (scholar) has the more likely he is deeply involved in a local network of collaboration (his research group) **(**Breschi and Lissoni, 2006). The higher degree of network clustering that is obtained by averaging C_i over all nodes in the system shows the existence of more cliques (closed research groups) within the network.

We have considered the top 5% with the highest clustering coefficient as the researchers most willing to collaborate with each other and create thus more cliques and act as embedded scientists in our study.

Author	Nano-Articles	Citation Count	H-index	Location	Clustering Coefficient
Summer Syed	2	31	1	Hamilton	**1**
R. Bader	13	4825	28	Hamilton	**1**
Suresh Tikoo	1	1295	25	Saskatoon	**1**
Leonard Fosteer	1	0	1	Vancouver	**1**
M. Kwok	1	7	2	Vancouver	**1**
Rajavel Elango	13	151	7	Vancouver	**1**
Andrea Damascelli	5	61	3	Vancouver	**1**
Hanane Becha	2	152	2	Ottawa	**1**
Amrutlal Patel	1	159	2	Saskatoon	**1**
Valerie Centis	1	135	4	Sherbrooke	**1**

Table 10: The top ten researchers with highest clustering coefficient

In the present study, we want to find out if the groupings that were proposed were significantly different from the general population in terms of average research performance (number of nanotechnology articles, citation count, and H-index). Henceforth we will use pairwise comparison hypothesis testing to justify our grouping position.

A statistical hypothesis test is a method of making decisions using data from a study. In statistics, a result is called statistically significant if it has been predicted as unlikely to have occurred by chance alone, according to a pre-determined threshold probability, the significance level. These tests are used in determining what outcomes of a study would lead to a rejection of the null hypothesis for a pre-specified level of significance; this can help to decide whether results contain enough information to cast doubt on conventional wisdom, given that conventional wisdom has been used to establish the null hypothesis.

The methodology used here was to remove the group data from the complete database, then using simple random sampling to collect a sample of approximately 5% of the complete data set. We can then use the two samples in our different hypothesis testing of means with known variances and assume independence between the samples. The Null Hypothesis for two means is $H_0 : \mu_i = \mu_j$ and $\sigma 1$ and $\sigma 2$ are Known.

We have tested the two-sided hypothesis at significant level of 0.05 and present the results on differences of means of our variable of interest for each group. The general results of our hypothesis testing decisions show that there is a significant statistical difference between the performance of scientists in each proposed group and the population. In other words the scholars belonging to each group are behaving differently than the population in terms of number of published articles, citations counts and h-index.

Groups	Nano Articles	Citations count	H-index
Star Scientists	Reject H0	Reject H0	Reject H0
Gatekeepers	Reject H0	Reject H0	Reject H0
Popular Scientists	Reject H0	Reject H0	Reject H0
Loyal Scientists	Accept H0*	Reject H0	Reject H0
Embedded Scientists	Reject H0	Reject H0	Reject H0

Table 11: Summary of decisions for each significance test at 95% confidence level

Although we could not reject the null hypothesis for the difference in means of publication for loyal scientists, but we rejected the two other tests regarding the citations count and H-index. Thus, we can also conclude that the research performance of loyal scientists group is different than the population.

5.3 Research Performance and Collaboration Behavior Analysis

In this section we present the analysis of our data sets using Exploratory Data Analysis (EDA) to summarize their main characteristics. Primarily EDA is used for seeing what the data can tell us beyond the formal modeling or hypothesis-testing task. Exploratory data analysis was promoted by John Tukey (1977) to encourage statisticians to explore the data, and possibly formulate hypotheses that could lead to new data collection and experiments.

The tables below present the simple statistics of the research performance indicators and network properties for each group.

	N	Mode	Mean	Std Dev	Sum	Min	Max
Nano Articles	679	10	23.01	20.31	15621	10	236
All Articles	679	34	98.76	87.83	67056	11	492
Citation Count	679	261	1358	1851	922097	9	26381
H_index	679	10	17.04	10.13	11568	1	81
Betweenness	679	0	0.00	0.00	0.39	0	0.02
Degree	679	1	4.33	5.94	2943	1	54
Weighted degree	679	1	2.04	1.60	1388	1	16
Clustering Coefficient	679	0	0.09	0.20	62.83	0	1
Strength of tie	679	1	5.48	9.72	3668	1	90

Table 12: EDA for Star scientists

	N	Mode	Mean	Std Dev	Sum	Min	Max
Nano Articles	682	1	6.96	17.19	4749	1	236
All Articles	682	5	70.53	82.55	48104	1	412
Citation Count	682	1	1099	1840	749703	1	17134
H_index	682	2	13.02	11.22	8882	1	67
Betweenness	682	0.002	0.00	0.00	2.36	0.00	0.04
Degree	682	7	11.49	10.57	7835	2	125
Weighted degree	682	1	2.39	1.58	1632	1	16.89
Clustering Coefficient	682	0	0.10	0.14	69.93	0	0.95
Strength of tie	682	1	4.83	7.84	8537	1	121

Table 13: EDA for Gatekeepers

	N	Mode	Mean	Std Dev	Sum	Min	Max
Nano Articles	685	1	5.42	16.00	3715	1	236
All Articles	685	5	47.98	73.03	32866	1	412
Citation Count	685	2	733.49	1578	502443	0	17134
H_index	685	1	9.80	10.11	6716	1	67
Betweenness	685	0	0.00	0.00	1.44	0	0.04
Degree	685	10	16.76	10.05	11480	10	125
Weighted degree	685	1	2.35	1.62	1608	1	16.89
Clustering Coefficient	685	0	0.32	0.29	217.57	0	1
Strength of tie	685	1	5.24	8.56	14714	1	121

Table 14: EDA for Popular scientists

	N	Mode	Mean	Std Dev	Sum	Min	Max
Nano Articles	688	1	3.75	6.77	2583	1	84
All Articles	688	1	38.81	58.01	26704	1	373
Citation Count	688	1	583.42	1239	401396	0	17134
H_index	688	1	9.16	8.68	6302	1	62
Betweenness	688	0	0.00	0.00	0.48	0	0.04
Degree	688	1	5.16	8.73	3553	1	125
Weighted degree	688	4	6.67	3.10	4591	4	31
Clustering Coefficient	688	0	0.14	0.22	97.79	0	1
Strength of tie	688	2	6.36	9.41	13327	1	121

Table 15: EDA for Loyal scientists

	N	Mode	Mean	Std Dev	Sum	Min	Max
Nano Articles	689	1	2.18	3.52	1505	1	58
All Articles	689	1	16.16	31.65	11136	1	248
Citation Count	689	1	256.86	710.73	176974	0	7046
H_index	689	1	4.98	6.16	3431	1	40
Betweenness	689	0	0.00	0.00	0.02	0	0.00
Degree	689	2	6.00	6.70	4132	2	46
Weighted degree	689	1	1.64	1.27	1133	1	17.8
Clustering Coefficient	689	1	0.95	0.09	651.59	0.74	1
Strength of tie	689	2	3.60	3.02	4346	1	17

Table 16: EDA foe Embedded scientists

For representing the distribution of several groups of variables associated with the scientists in each category around the mean, we used the boxplot. Note that the highest datum represents only 25% of the upper quartile due to the large amount of data distributed above the 75th percentile. The following sections show the result of analyzing the distribution of research performance indictors' data followed by the distribution of collaboration behaviour characteristics.

- **Research Performance Data**

For the number of articles, we excluded the data for star scientists since they were selected based on this variable. The distribution of number of publications in each group shows that the gatekeepers and popular scientists have the highest number of publications where the majority of the data are above the median and they have the highest maximum number of publications comparing to the others.

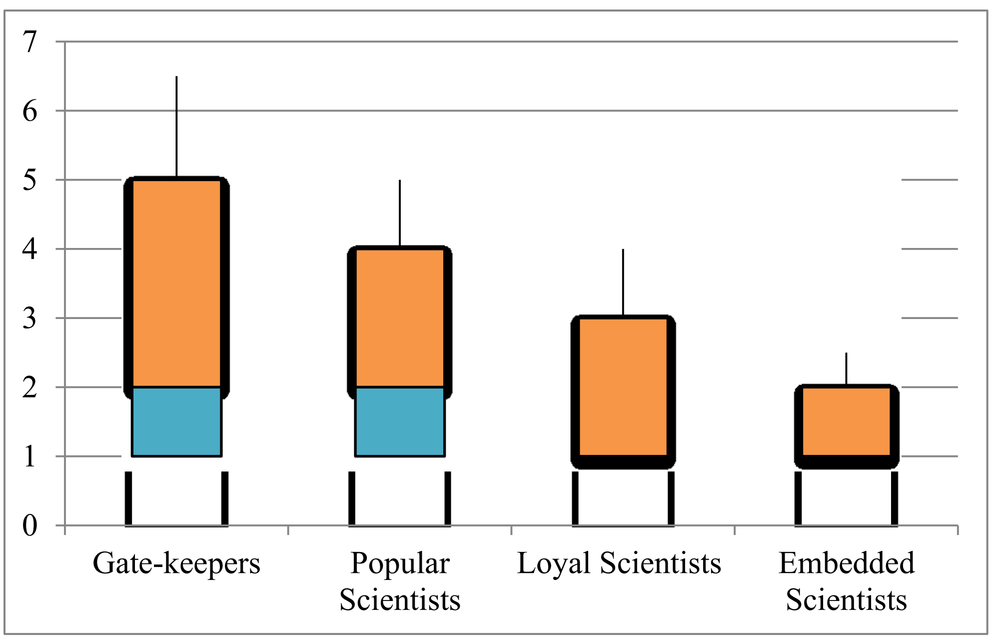

Figure 13: Boxplot for the number of nanotechnology articles per group

In terms of citations count, star scientists followed by gatekeepers, as expected, have the highest citations count. Therefore, the majority of the scholars belonging to these groups have higher citations count than the median. On the other hand, both popular scientists and loyal ones behave similarly with lower performance than the previous mentioned ones. In addition, embedded scientists showed the worst performance in terms of citations count. Scholars in this group have the lowest minimum and large amount of data is distributed under the median.

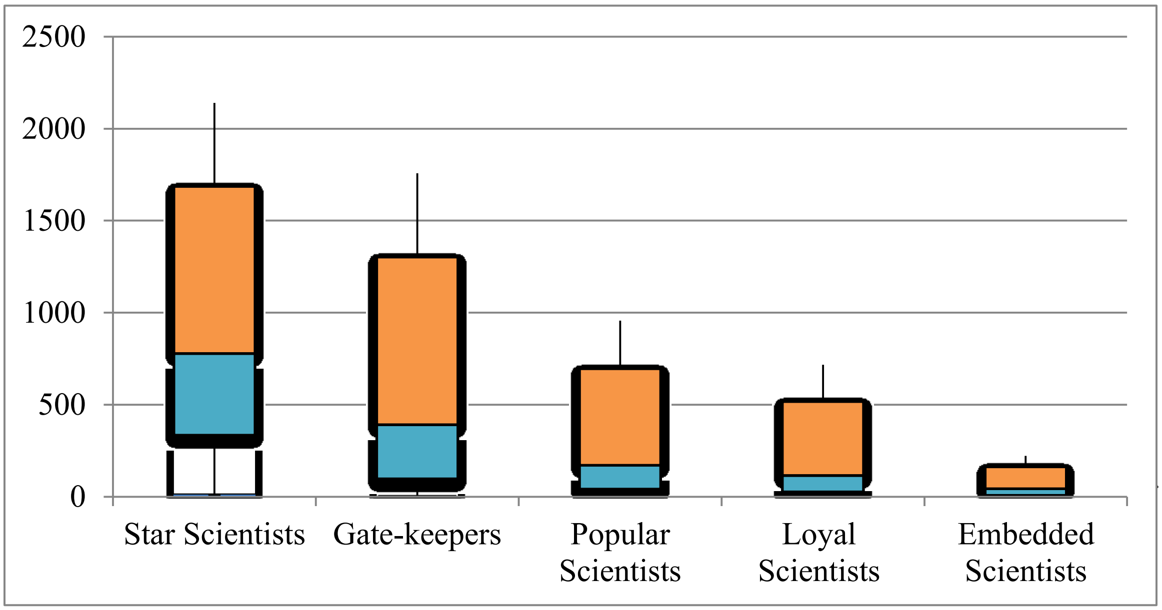

Figure 14: Boxplot for the citations count per group

Regarding the H-index data, star scientists and gatekeepers also have the best performance comparing to others. Furthermore, embedded scientists have the lowest performance while the rest two groups are performing in the average.

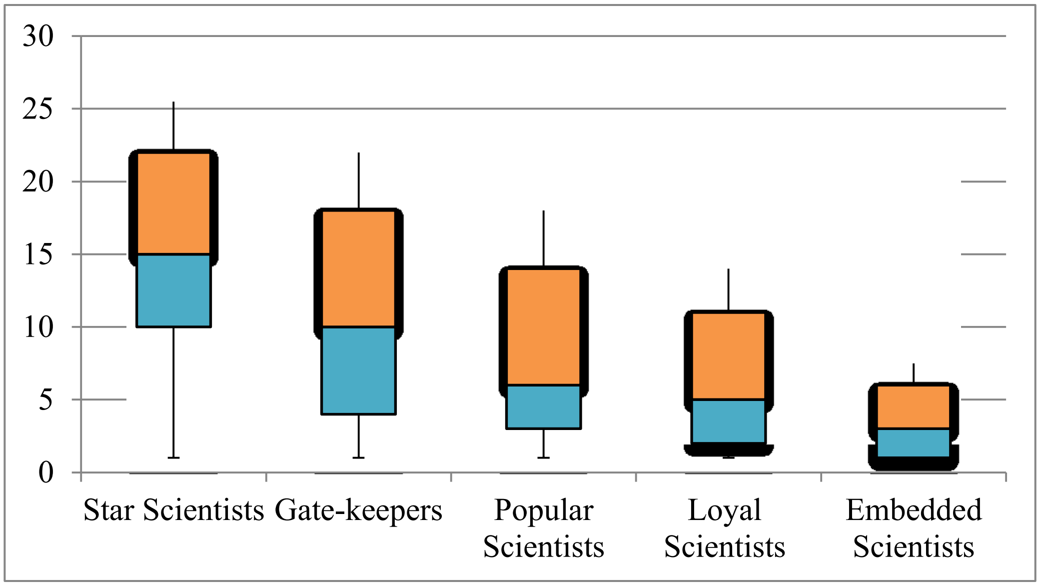

Figure 15: Boxplot for the H-index per group

- **Collaboration Behaviour Data**

As mentioned earlier, we measured the strength of collaboration tie between two scholars by the number of repeated collaboration experience they have together. Loyal scientists mainly repeat the collaboration with the same group of partners and consequently have the strongest collaboration ties. Referring to the previous performance analysis which showed bad performance of loyal scientists (comparing to the other groups), we can say that having single collaboration with many partners is better than maintaining the same collaboration relationships. Conversely, star scientists, who perform the best, have the weakest collaborations ties, which mean that collaborating with new partners positively affects the process of knowledge creation and sharing.

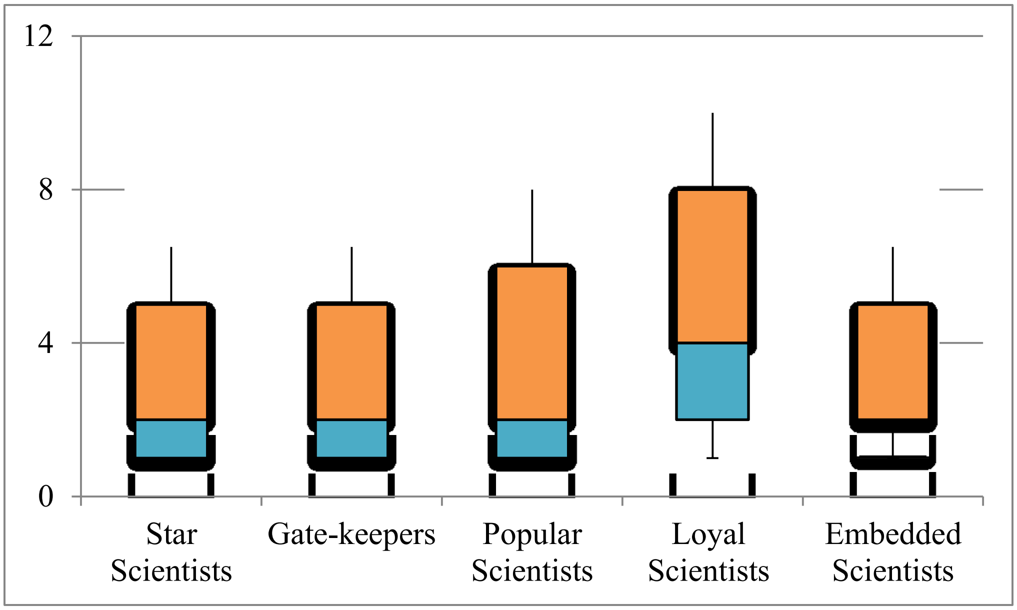

Figure 16: Boxplot for the strength of collaboration ties per group

We have analyzed the distribution of scientists belonging to the defined group based on other qualitative data such as their location, their affiliation and primary affiliation of the preferred collaborators. The histogram below graphically represents the distribution of locations of scientists in each group. The figure shows the distribution of each group in the six cities that have the majority of the scholars.

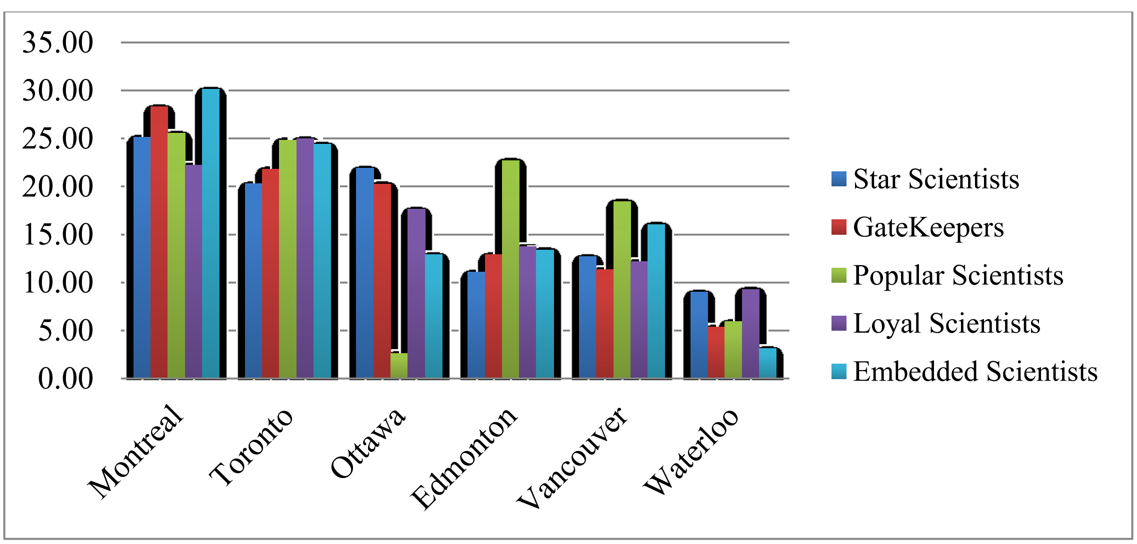

Figure 17:Histogram for the distribution of scientists' location per group

The table below summarizes the quantitative and qualitative data regarding the patterns of research performance and collaboration behaviour for each group.

	Star Scientists		Gate-Keepers		Popular Scientists		Loyal Scientist		Embedded Scientists	
	Mode	**Avg.**	**Mode**	**Avg.**	**Mode**	**Avg.**	**Mode**	**Avg.**	**Mode**	**Avg.**
Nano Articles	10	**23**	1	**7**	1	**5**	1	**4**	1	**2**
All articles	34	**99**	5	**71**	5	**48**	1	**39**	1	**16**
Citation Count	261	**1358**	1	**1099**	1	**733**	1	**538**	1	**257**
H-index	10	**17**	2	**13**	1	**10**	1	**9**	1	**5**
Strength of tie	1	**4**	1	**5**	1	**5**	2	**7**	2	**4**
Location	Montreal		Montreal		Montreal		Toronto		Montreal	
Affiliation	Academia		Academia		Academia		Academia		Academia	
Preferred Collaborators	Academia		Academia		Academia		Academia		Academia	

Table 17: Summary of research performance and collaboration behavior pattern for each group of scientists

5.4 Partners' Selection Mechanism

The response rate for this survey was unexpectedly high 20%. A total of 281 scientists (235 male, 84%; and 46 female, 16%) ranging between so called beginners, and those of intermediate and advanced levels of research experience were recruited for this study. The largest portion of participants 75% was those who are affiliated to academic organizations in different positions; graduate students, postdoctoral fellows, research assistants, assistant professors, associate professors and professors. The respondents belonged to different fields of expertise; biology, chemistry, physics, engineering and medicine; and they are distributed all over the Canadian provinces with 5% who are currently residing outside Canada.

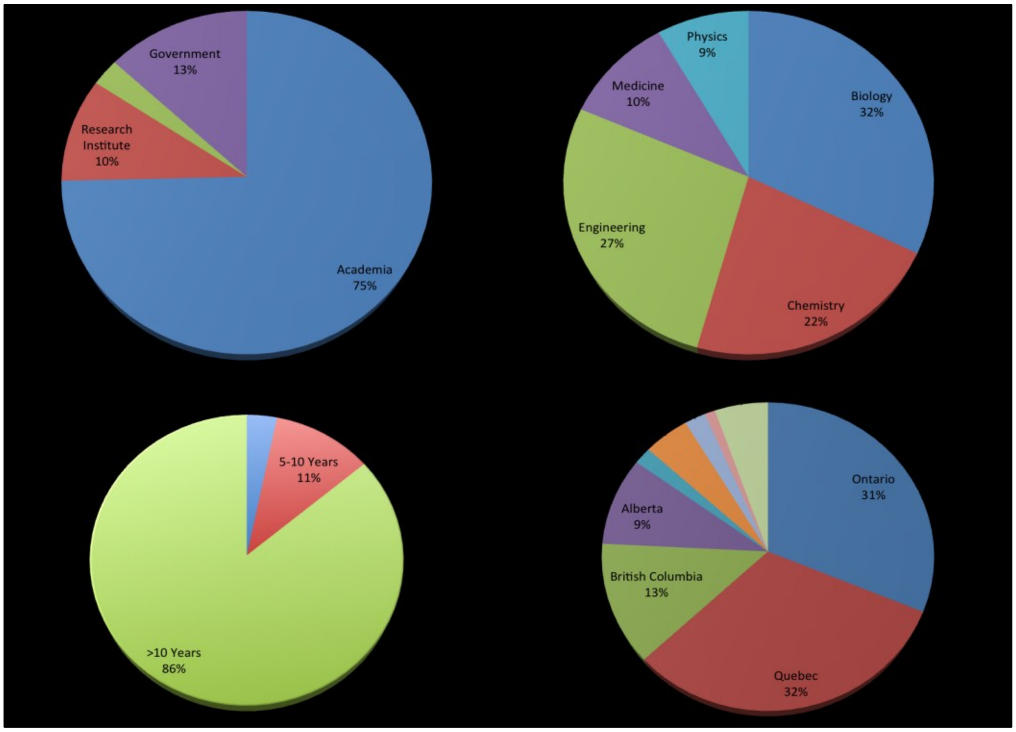

Figure 18: The percentage of participants in the questionnaire

The figure above shows the percentage of participants who responses to the questionnaire based on (i) the category of the organization they are affiliated to, (ii) their primary field of expertise, (iii) the number of experience years in research they have, and (iv) their current residence location.

For analyzing the responses, we have assigned weights for each answer in the Likert scale as: Unimportant (-2), Slightly Important (-1), Important (0), Very Important (1), and Critically Important (2). The Likert-scale questions included the following factors:

F1-- The reputation of the organization that research partner is currently affiliated to

F2-- The accessibility to resources, required tools and equipment in research partner's organization

F3-- The availability of funding the research partner is bringing to the project

F4-- Research partner's total number of publications and patents

F5-- Research partner's publications citation rate

F6-- Research partner's career age and years of research experience

F7-- Research partner's reputation in the field i.e. the authors' H-index, citation count and being well known in the field

F8-- Research partner has common research interest

F9-- Research partner's knowledge in complementary field(s)

F10-- Research partner's personal relationship to you, i.e. friends and family

F11-- Research partner is already within your professional network

F12-- Research partner's prior satisfactory collaboration experience with you

F13-- The strength of the collaboration tie, i.e. the number of your previous common projects and/or the duration of collaboration's relation

F14-- Research partner's geographical location

F15-- Research partner's native language

F16-- Research partner's cultural background

F17-- Research partner's gender

F18-- Research partner's age

The table below presents the descriptive analysis for the answers in each choice:

	Unimportant	Slightly Important	Important	Very Important	Critically Important	# of Respondents	Rating Average
F1	17% (47)	**32% (90)**	28% (78)	16% (44)	7% (20)	279	-0.36
F2	6% (16)	10% (27)	28% (80)	**33% (94)**	22% (62)	279	0.55
F3	7% (20)	18% (51)	**33% (94)**	27% (77)	13% (37)	279	0.21
F4	24% (68)	**36% (100)**	27% (75)	9% (26)	3% (8)	277	-0.69
F5	32% (89)	**34% (96)**	20% (57)	11% (30)	2% (6)	278	-0.83
F6	33% (93)	**36% (100)**	21% (60)	8% (22)	1% (4)	279	-0.92
F7	7% (21)	18% (50)	**39% (109)**	23% (66)	12% (33)	279	0.15
F8	5% (14)	13% (36)	26% (72)	**29% (82)**	26% (72)	276	0.58
F9	2% (6)	4% (12)	17% (48)	34% (96)	**40% (112)**	274	1.06
F10	**42% (119)**	27% (76)	15% (43)	8% (23)	5% (14)	275	-0.93
F11	30% (84)	**31% (88)**	25% (70)	10% (27)	4% (10)	279	-0.73
F12	5% (15)	11% (30)	30% (83)	**30% (84)**	22% (63)	275	0.59
F13	10% (27)	17% (49)	**37% (104)**	22% (62)	12% (35)	277	0.09
F14	**38% (108)**	28% (78)	20% (55)	12% (35)	0% (1)	277	-0.92
F15	**60% (169)**	25% (70)	7% (21)	4% (11)	1% (4)	275	-1.39
F16	**73% (204)**	19% (52)	6% (16)	1% (3)	0% (1)	276	-1.64
F17	**92% (258)**	6% (17)	0% (1)	0% (0)	0% (0)	276	-2.01
F18	**82% (230)**	15% (42)	1% (4)	0% (0)	0% (0)	276	-1.78

Table 18: Descriptive analysis. Summary of responses for each answer choice

Due the low response rate for some answer choices and for statistical analysis purpose, we have reduced the answers levels into 3 instead of 5. The lowest ranks choices (unimportant and slightly important) have been combined into NEGATIVE, the middle point (important) stated as NEUTRAL, where POSITIVE represents the highest ranks choices (very important or critically important).

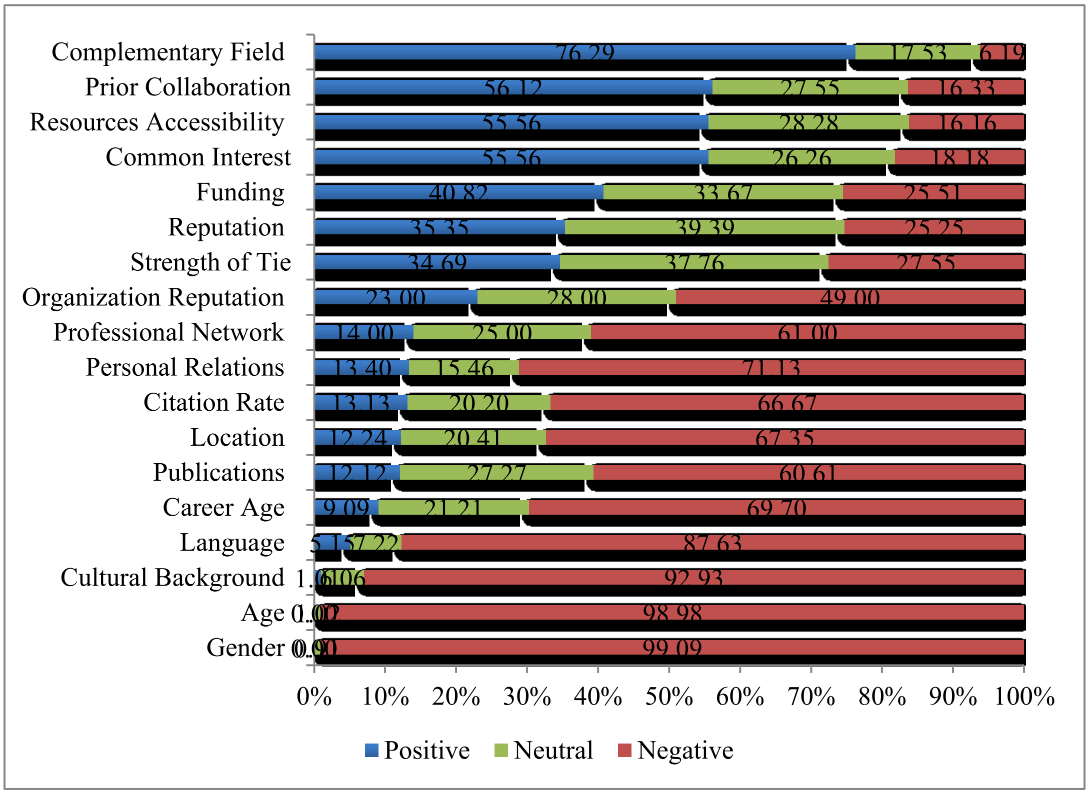

Figure 19: The percentage of total positive, neutral and negative answers

The previous table shows the percentage of respondents for each answer to all 18-items. It shows also the rating average based on the weight assigned to each answer choice. The rating average is calculated as follows, where:

w = weight of answer choice, x = response count for answer choice

$$\frac{x1w1 + x2w2 + x3w \ldots xnwn}{Total}$$

The rating average showed that 7 out of the 18 factors have a positive impact on the scientists' decision regarding their potential partners. F2, F3, F7, F8, F9, F12 and F13 are significantly affecting the partners' selection mechanism, where the personal factors F16, F17 and F18 have no impact on the partners' selection decision.

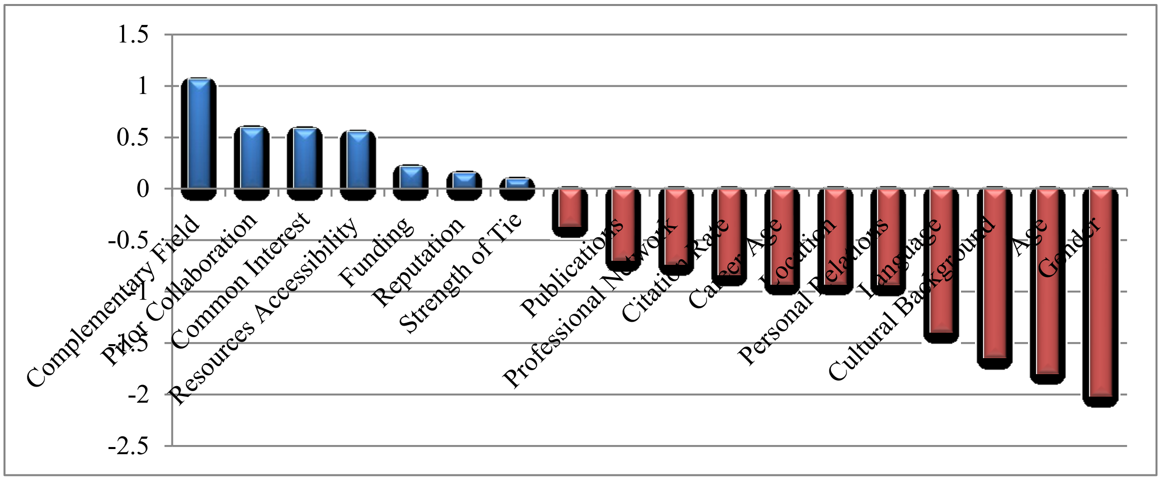

Figure 20: The partners' selection factors ordered by their importance level

5.4.1 Importance of Factors Based on Years of Experience

The importance level of the highest-ranking factors has been analyzed for several groups of the respondents based on how many years of research experience they have. We have here three groups, which are advanced researchers with more than 10 years of experience (N= 242, 86%), the intermediate researchers with 5 to 10 research experience years (N=30, 11%), and the beginners who have been engaged in research activities for less than 5 years (N=9, 3%).

Here we have analyzed the importance of each factor for each group:

Group I: Advanced researchers with more than 10 years of experience

Motivations	N	Rank Sum	Mean Rank
Resources Accessibility	239	142.5	0.60
Funding	238	57	0.24
Partner's Reputation	238	34.5	0.14
Common Interest	236	138	0.58
Complementary Field	233	256.5	**1.10**
Prior Collaboration	236	138	0.58
Strength of Tie	237	31.5	0.13

Table 19: The importance of the highly ranked factors for the advanced researchers

Research partner's knowledge in complementary field 1.10, followed by the accessibility to resources, required tools and equipment in research partner's organization 0.60 are indications of greater significance level in selecting the potential partner(s). While the reputation of the partner in the field 0.14 and the strength of the collaboration tie had the lowest mean rank 0.13 and thus lowest importance level for this group of scientists.

Group II: Intermediate researchers with 5 to10 years of experience

Motivations	N	Rank Sum	Mean Rank
Resources Accessibility	28	27	**0.96**
Funding	29	6	0.21
Partner's Reputation	29	9	0.31
Common Interest	28	12	0.43
Complementary Field	29	27	**0.93**
Prior Collaboration	28	16.5	0.59
Strength of Tie	29	4.5	0.16

Table 20: The importance of selection factors for the intermediate researchers

Both the accessibility to resources, required tools and equipment in research partner's organization 0.96 and research partner's knowledge in complementary field 0.93 have a great influence on the scientists' decision regarding selecting their partners. On the other hand, the availability of funding the research partner is bringing to the project 0.21 and the strength of the collaboration tie had the lowest mean rank 0.16 had the lowest mean rank 0.14.

Group III: Beginner researchers with less than 5 years of experience

Motivations	N	Rank Sum	Mean Rank
Resources Accessibility	8	-3	-0.38
Funding	6	-1.5	-0.25
Partner's Reputation	12	1.5	0.13
Common Interest	7	0	0.00
Complementary Field	10	1.5	**0.15**
Prior Collaboration	11	1.5	0.14
Strength of Tie	9	-3	-0.33

Table 21: The importance of selection factors for the intermediate researchers

Research partner's knowledge in complementary field 0.15, followed by the prior satisfactory collaboration experience 0.14, are indications of greater significance level in selecting the potential partner(s); where the lowest importance level is for resources accessibility with lowest mean rank -0.38 and, not surprisingly, followed by the strength of collaboration tie -0.33.

The low mean rank in this group, which depends on the number of answers compared to others, is indeed influenced by the low response rate we got from this group. As mentioned before only 3% of the respondents to our survey were actually beginners in the research field with less than 5 years of experience. In fact, most of those beginners in real world do not large number of publications, or probably do not have their articles in SCOPUS yet, thus their percentage in our population is low.

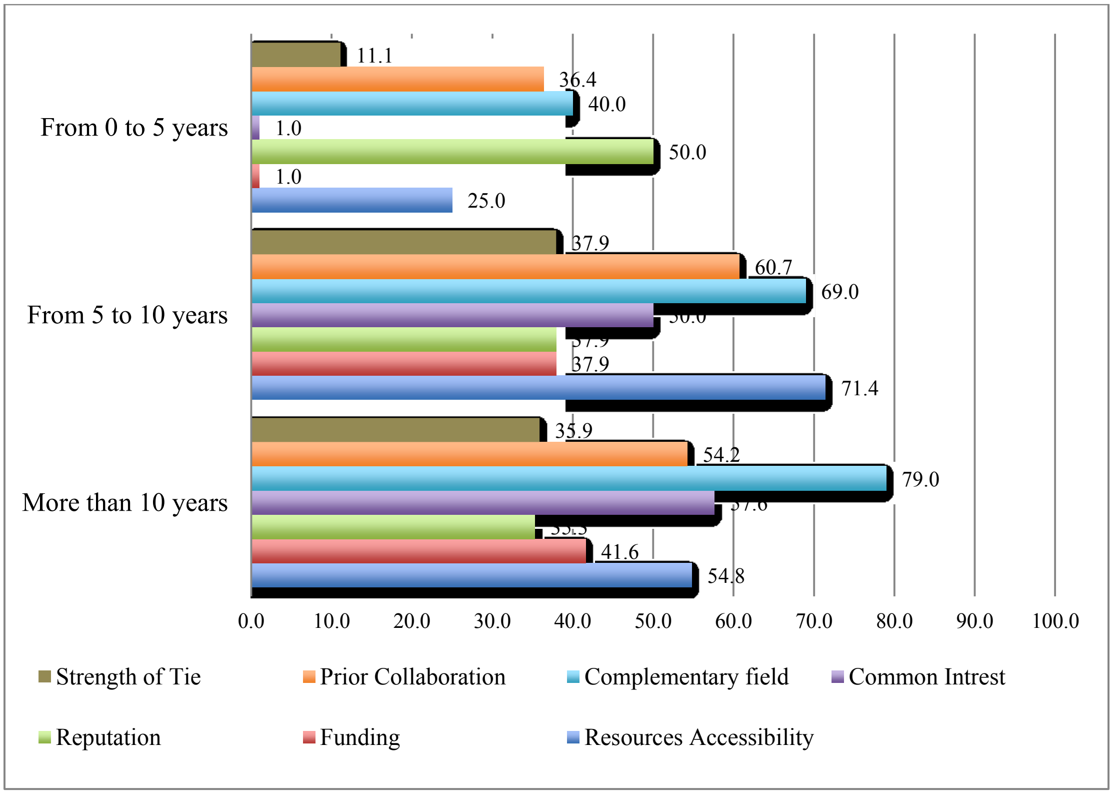

Figure 21: The percentage of positive answers based on research experience

Indeed, it was expected that the beginners would care mostly about funding and resource accessibility, and then later you care more about the reputation and prior satisfactory experience. Complementary field or common interest would be expected to rank similarly regardless their research experience. The unexpected results here show that beginners do put much weight on the funding and common interest when searching for research partners, but this can be also explained by the low response rate due the reasons disused earlier. On the other hand, finding a partner who is well known and has a high academic reputation shows a steady importance level for all groups of scientists and especially for those with minimal level of experience.

5.4.2 Importance of Factors Based on the Primary Affiliations

The importance level of the highest-ranking factors has been analyzed for several groups of the respondents based on their primary affiliations. The responses to the category of organization question were as following: academia (N= 206, 75%), government (N=37, 13%), research institutes (N=30, 10%), and industry (N=7, 2%). Some other answers include: hospital, non-profit organization, and consulting engineering agency.

Here we have analyzed the importance of each factor for the four main groups:

Group I: Researchers affiliated to academic institutions

Motivations	N	Rank Sum	Mean Rank
Resources Accessibility	201	111	0.55
Funding	201	30	0.15
Partner's Reputation	201	13.5	0.07
Common Interest	199	121.5	0.61
Complementary Field	197	208.5	**1.06**
Prior Collaboration	198	108	0.55
Strength of Tie	199	13.5	0.07

Table 22: The importance of the selection factors for researchers in academia

Research partner's knowledge in complementary field 1.06, followed by both the accessibility to resources, required tools and equipment in research partner's organization and the prior satisfactory collaboration experience 0.55 are indications of greater significance level in selecting the potential partner(s). While the reputation of the partner in the field and the strength of the collaboration tie had the lowest mean rank 0.07 and thus lowest importance level for this group of scientists.

Group II: Researchers affiliated to governmental organizations

Motivations	N	Rank Sum	Mean Rank
Resources Accessibility	37	27	0.73
Funding	37	19.5	0.53
Partner's Reputation	37	12	0.32
Common Interest	37	15	0.41
Complementary Field	37	42	**1.14**
Prior Collaboration	37	31.5	0.85
Strength of Tie	37	12	0.32

Table 23: The importance of selection factors for researchers in governmental agencies

Research partner's knowledge in complementary field 1.14, followed by the prior satisfactory collaboration experience 0.85 is indication of greater significance level in selecting the potential partner(s). While the reputation of the partner in the field and the strength of the collaboration tie had the lowest mean rank 0.32 and thus lowest importance level for this group of scientists.

Group III: Researchers affiliated to research institutes

Motivations	N	Rank Sum	Mean Rank
Resources Accessibility	28	21	0.75
Funding	28	10.5	0.38
Partner's Reputation	28	9	0.32
Common Interest	27	9	0.33
Complementary Field	27	24	**0.89**
Prior Collaboration	28	7.5	0.27
Strength of Tie	28	7.5	0.27

Table 24: The importance of selection factors for researchers in research institutes

Research partner's knowledge in complementary field 0.89, followed by accessibility to resources, required tools and equipment in research partner's organization 0.75 are indications of greater significance level in selecting the potential partner(s). While both the prior satisfactory collaboration experience and the strength of the collaboration tie had the lowest mean rank 0.27 and thus lowest importance level for this group of scientists.

Group IV: Researchers affiliated to industry

Motivations	N	Rank Sum	Mean Rank
Resources Accessibility	6	4.5	0.75
Funding	6	3	0.50
Partner's Reputation	6	1.5	0.25
Common Interest	4	1.5	0.38
Complementary Field	3	3	**1.00**
Prior Collaboration	7	1.5	0.21
Strength of Tie	7	-1.5	-0.21

Table 25: The importance of selection factors for researchers in industry

Research partner's knowledge in complementary field 1.00, followed by accessibility to resources, required tools and equipment in research partner's organization 0.75 are indications of greater significance level in selecting the potential partner(s). While both the prior satisfactory collaboration experience 0.21 and the strength of the collaboration tie -0.21 had the lowest mean rank and thus lowest importance level for this group of scientists.

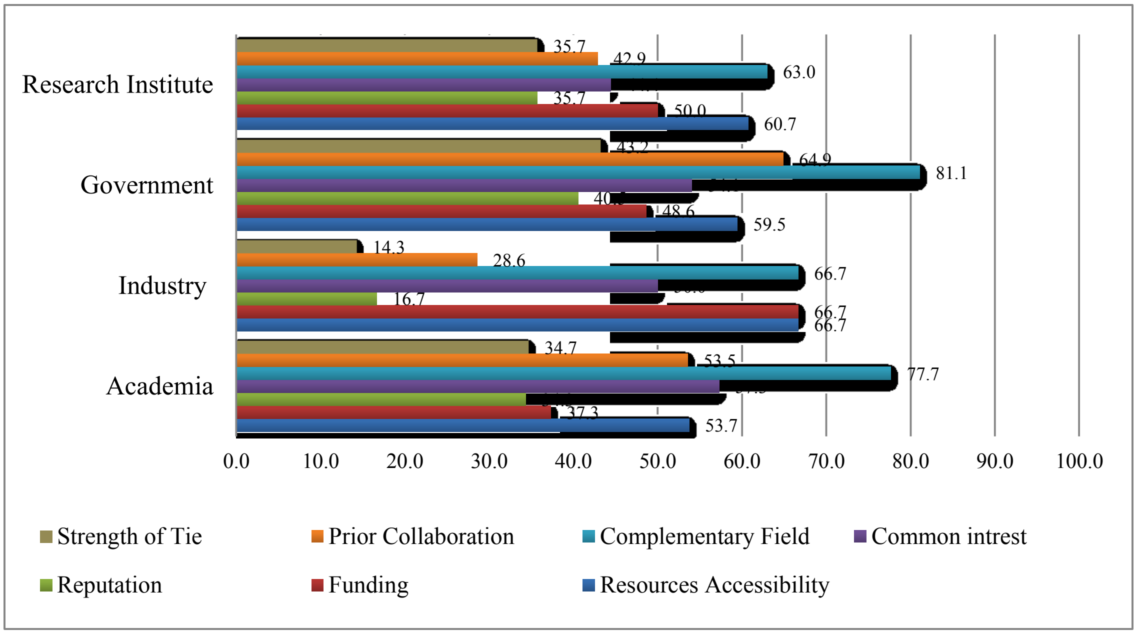

Figure 22: The percentage of positive answers based on primary affiliations

Obviously, complementary field is very important for everybody, but it was expected that academicians search more for funding, and our result shows that it is fact industry, which is in need of money. We can therefore suggest that funding is an important factor to be considered while seeking a research partner. Even in the case were a budget is allocated for a research project, i.e. for researchers work for industry, sometimes it would be cheaper for researchers to collaborate and share the experimental expenses.

Moreover, it is probably expected that industrial researchers do not consider the reputation as very important, as they need more practical results applicable for their company than publications, citations, scientific accolades and fame. Whereas collaborating with scientists from complementary field is of a high importance for almost all scientists. The integration of their ideas with those of the researchers from a dissimilar field is specifically critically in a multidisciplinary sector such as nontechnology.

5.4.3 Importance of Factors Based on the Field of Expertise

The importance level of the highest-ranking factors has been analyzed for several groups of the respondents based on their primary field of expertise. We categorize the scientists based on their primary major according to the fields' categorization in academic software called 'Publish or Perish[28]'. The respondents to our survey belonged to the different following majors: Biology, Life Sciences, Environmental Science (N= 89, 32%), Chemistry and Materials Science (N=61, 22%), Engineering, Computer Science, and Mathematics (N=75, 27%), Medicine, Pharmacology, and Veterinary Science (N=28, 10%), and Physics, Astronomy, and Planetary Science (N=23, 9%).

Here we have analyzed the importance of factor for each group:

Group I: Researchers in Biology, Life Sciences, and Environmental Science

Motivations	N	Rank Sum	Mean Rank
Resources Accessibility	89	69	0.78
Funding	89	33	0.37
Partner's Reputation	89	19.5	0.22
Common Interest	88	51	0.58
Complementary Field	88	136	**1.55**
Prior Collaboration	88	54	0.61
Strength of Tie	88	3	0.03

Table 26: The importance of selection factors for researchers in Biology, Life Sciences, and Environmental Science

Research partner's knowledge in complementary field 1.55, followed by the accessibility to resources, required tools and equipment in research partner's organization 0.78 are indications of greater significance level in selecting the potential

[28] Publish or Perish is a software program that retrieves and analyzes academic citations using Google Scholar, *available from http://www.harzing.com/pop.htm*

partner(s). While both the reputation of the partner in the field 0.22 and the strength of the collaboration tie 0.03 had the lowest mean rank and thus lowest importance level for this group of scientists.

Group II: Researchers in Chemistry and Materials Science

Motivations	N	Rank Sum	Mean Rank
Resources Accessibility	61	37.5	0.61
Funding	61	-3	-0.05
Partner's Reputation	59	7.5	0.13
Common Interest	58	27	0.47
Complementary Field	58	63	**1.09**
Prior Collaboration	59	37.5	0.64
Strength of Tie	61	37.5	0.42

Table 27: The importance of selection factors for researchers in Chemistry and Material Science

Research partner's knowledge in complementary field 1.09, followed by the prior collaboration experience 0.64 is indication of greater significance level in selecting the potential partner(s). While the reputation of the partner in the field 0.13 and the availability of funding -0.05, had the lowest mean rank and thus lowest importance level.

Group III: Researchers in Engineering, Computer Science, and Mathematics

Motivations	N	Rank Sum	Mean Rank
Resources Accessibility	72	30	0.42
Funding	71	19.5	0.27
Partner's Reputation	73	7.5	0.10
Common Interest	73	45	0.63
Complementary Field	72	63	**0.88**
Prior Collaboration	72	34.5	0.49
Strength of Tie	71	3	0.04

Table 28: The importance of selection factors for researchers in Engineering, Computer Science, and Mathematics

Research partner's knowledge in complementary field 0.88, followed by the common research interest 0.63 is indication of greater significance level in selecting the potential partner(s). While the reputation of the partner in the field 0.10 and the strength of collaboration tie 0.04, had the lowest mean rank and thus lowest importance level.

Group IV: Researchers in Medicine, Pharmacology, and Veterinary Science

Motivations	N	Rank Sum	Mean Rank
Resources Accessibility	28	28.5	1.02
Funding	28	10.5	0.38
Partner's Reputation	28	3	0.11
Common Interest	28	12	0.43
Complementary Field	27	28.5	**1.06**
Prior Collaboration	28	12	0.43
Strength of Tie	27	1.5	0.06

Table 29: The importance of selection factors for researchers in Medicine, Pharmacology, and Veterinary Science

Research partner's knowledge in complementary field 1.06, followed by the accessibility to resources, required tools and equipment in research partner's organization 1.02 are indications of greater significance level in selecting the potential partner(s). While the reputation of the partner in the field 0.11 and the strength of collaboration tie 0.06, had the lowest mean rank and thus lowest importance level.

Group V: Researchers in Physics, Astronomy, and Planetary Science

Motivations	N	Rank Sum	Mean Rank
Resources Accessibility	21	6	0.29
Funding	22	4.5	0.20
Partner's Reputation	22	7.5	0.34
Common Interest	22	15	0.68
Complementary Field	22	27	**1.23**
Prior Collaboration	22	13.5	0.61
Strength of Tie	22	-3	-0.14

Table 30: The importance of selection factors for researchers in Physics, Astronomy, and Planetary Science

Research partner's knowledge in complementary field 1.23, followed by the common research interest 0.68 is indication of greater significance level in selecting the potential partner(s). While the availability of funding 0.20 and the strength of collaboration tie -0.14, had the lowest mean rank and thus lowest importance level.

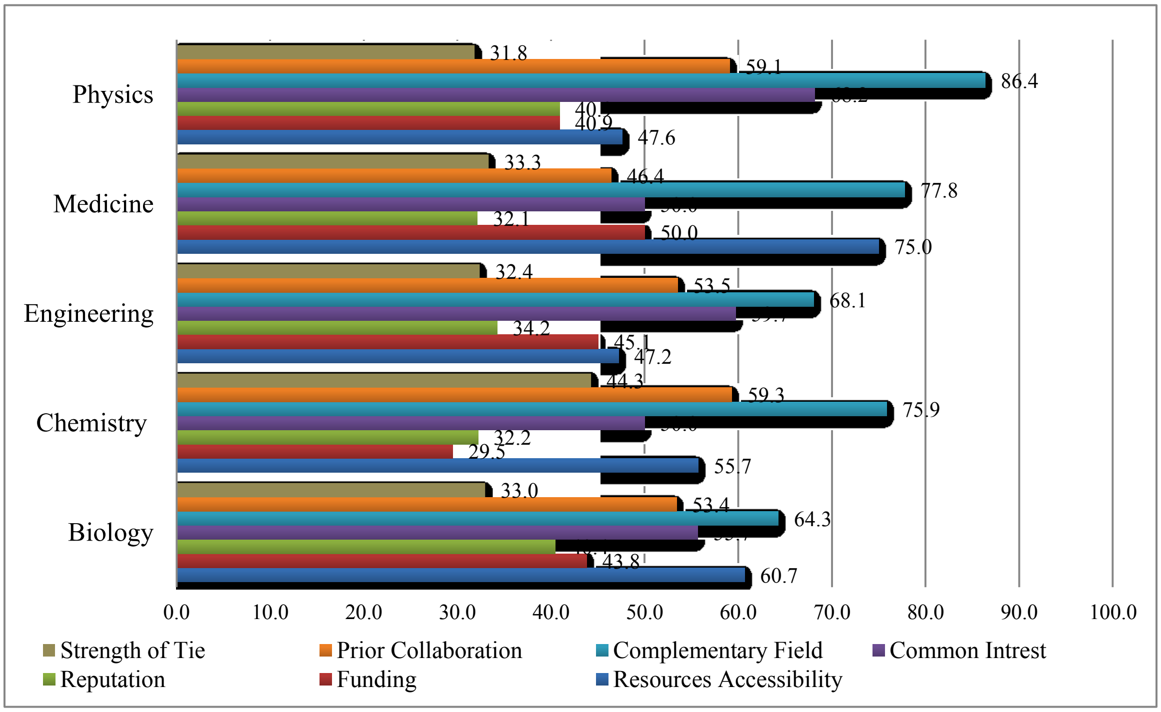

Figure 23: The percentage of positive answers based on field of expertise

Having a satisfactory collaboration experience, as expected showed a critical impact on influencing the scientists' decision regarding whom they collaborate with for all researchers regarding their field of expertise. Besides, common interest and complementary field are ranking much higher in physics. On the other hand, scientists in chemistry are the least to care about funding whereas those in medicine consider funding as more critical. Although there is a substantial research funding for medicine in Canada (at least more than for engineering), one would expect that money or resources are not that decisive factors for medicine. The results need more investigation for clarifying the reasons behind the different level of importance of these factors to scientists in each field.

We can generally conclude that our results coincided with some of the collaboration motivations that were mentioned in the literature. That is, funding, resource accessibility and the strong collaboration ties have been always the most critical consideration (for example Mat *et al.*, 2009 and Melin 2000). Furthermore, the collaboration with a partner from different field of expertise showed an increasing level of importance in nanotechnology due the fact that it is a multidisciplinary industry where scientists from different sciences and engineering fields are interested in.

On the other hand, although geographical location, nationalistic and personal relationships appeared to have significant correlation with the scientific collaboration in the literature (Lee and Bozeman, 2005 and Bozeman and Corley, 2004 and others), our results showed that these factors have no impact for the making the partnership's decision while seeking collaborators.

5.4 The Impact of Individual Scientists' behavior on the Knowledge Flows and Scientific Production

In the next sections, the performance and structure of innovation networks will be numerically analyzed for different scenarios, and the results always compared to the basic one that exemplify the real world where all groups are included with a 0.05 ratio to the population.

5.4.1 The Role of Star Scientists

The star scientists, according to the definition in the literature and in this work, are the main producers of scientific knowledge in the network. Moreover, according to the result of our survey they are more likely to be selected as partners due to their high academic reputation. Correspondingly, it is expected that their presence and absence will have different impact on the flows of knowledge in the information based innovation networks. In this section, we run our simulation model while removing the star scientists completely from the network to examine how their absence affects the behavior of other scientists comparing to the original scenario (where all groups are there including star scientists).

First, the performance of Canadian nanotechnology network is analyzed in the presence and absence of star scientists group. The figure below illustrate corresponding results of the average productivity of the scientists (i.e. total number of articles divided by the population of scientists) is both scenarios. The figure shows that the average productivity of the network is almost 25% less than its amount in the absence of star scientists in the network. The first scenario, where the star scientists are included, the average productivity of scientists in the network is 1.66 article/author, which is considerably higher than the average of 1.26 article/author in the second scenario when they were excluded.

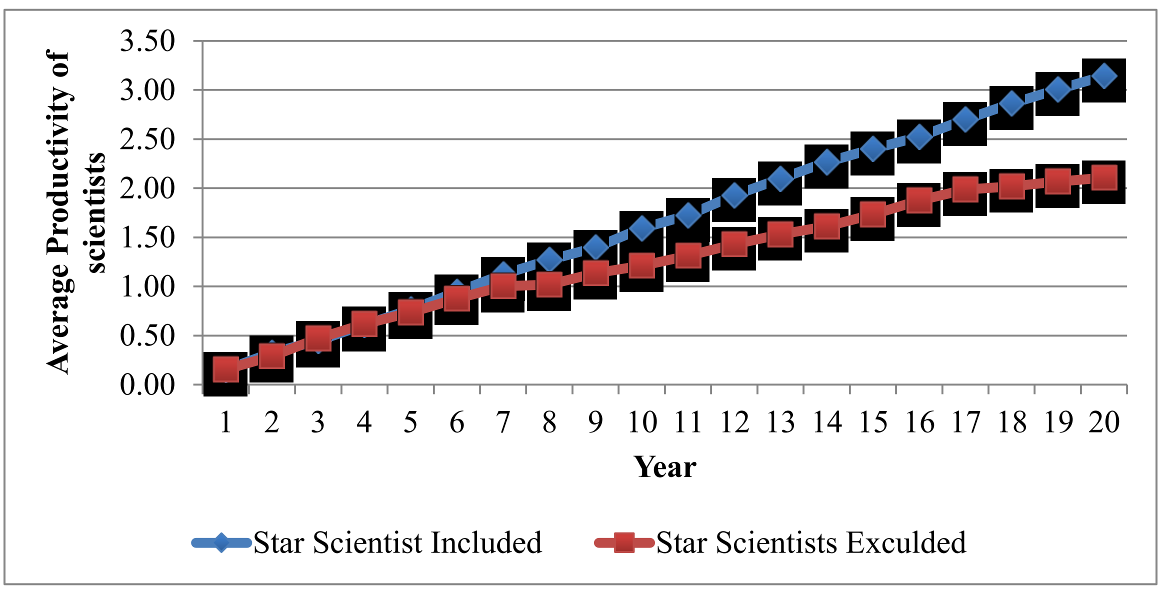

Figure 24:Average number of publications per scientist-Star scientists included/excluded

As star scientists are usually more attractive for partnership and can be selected by more than one scientist at the same time, that will result in increase of the average number of articles coauthored by each scientist in the network. That is, the absence of star scientists will give similar opportunities for all the scientists to be selected as partners, and therefore the share of productivity is more evenly distributed among the scientists.

Furthermore, the average number of publications over the 20 years reduces from 3436.35 publications/year with star scientists in the network to about 2951.62 publications/year in the case of their absence. The figure below displays the average number of publications in both scenarios for the ten replications. The figure shows that the productivity of the network is negatively affected by the absence of star scientists.

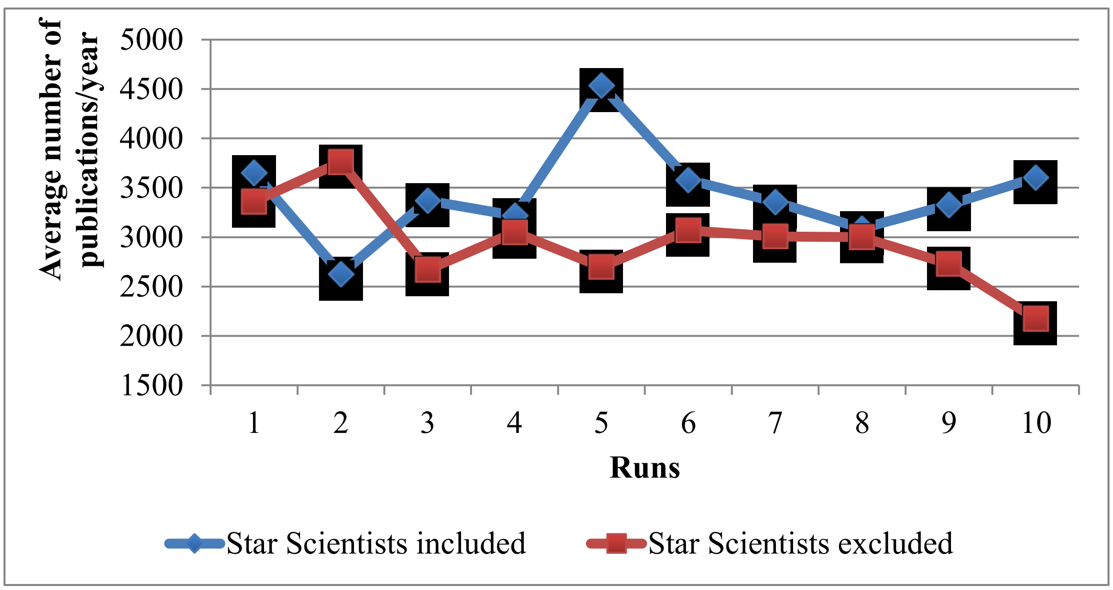

Figure 25: Average network productivity-Star scientists included/excluded

The less number of star scientists would give them higher chance to be selected over and over again and consequently enhance their individual performance and that positively contributes toward improving the overall productivity. The figure below shows the average productivity per scholar and the network performance in 5 different scenarios, which involve various percentages of stars to the population. The lower number of star scientists within the scientists' population in fact shows a better performance of the network, and with an increasing number of the star scientists in the network the average productivity decreases. This result is rather surprising as it was expected that having more star scientists in the network will increase the productivity of the system, and if there is any "optimal" number of star scientists it would be detected at a higher percentage. Our result can be explained by the network properties. If there are only very few individuals performing extremely well, the network becomes much more centralized which improves its overall knowledge transmission properties, and consequently has a positive impact on its performance. More research is needed to shed more light on this issue.

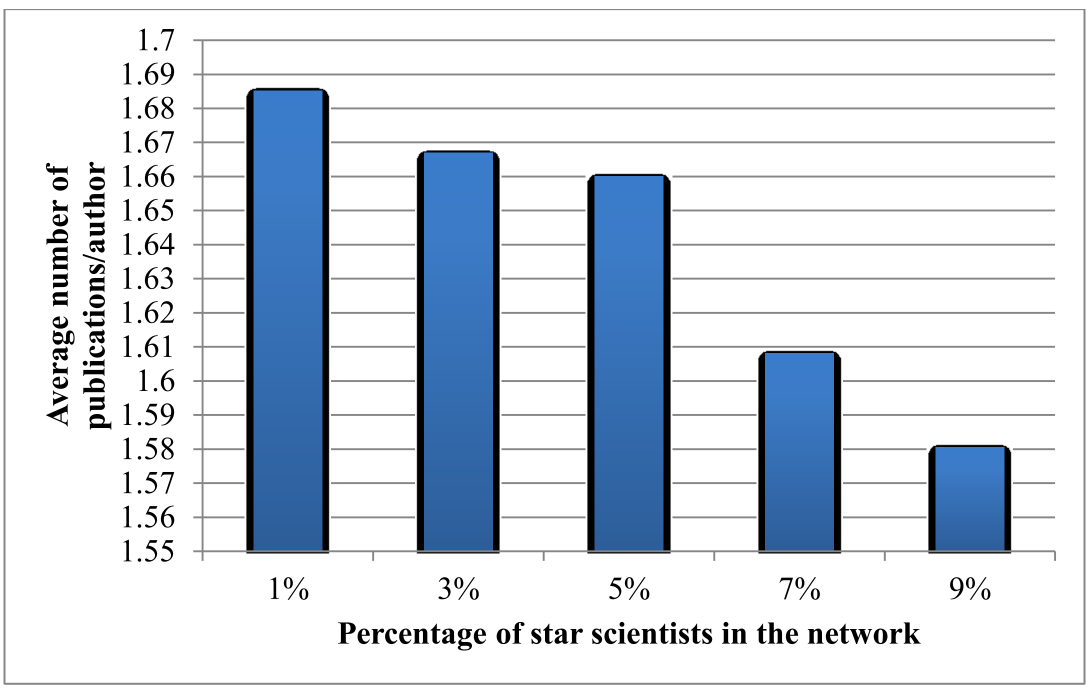

Figure 26: The average research performance with different percentage of star scientists

On the other hand, we have analyzed the performance of the other groups of scholars in case of excluding the star scientists. Star scientists usually occupy more central positions in the network, with higher number of connections (Schiffauerova and Beaudry, 2011). That is, the star scientists are also part of gatekeepers and popular scientists groups who have highest betweenness centrality and more connections. Consequently, removing them from the network will also affect the performance of these groups negatively. The table below shows the overall performance of each group of scientists when the star scientists exist and when they are absent. The performance is measured by the average number of articles coauthored by scientists belonging to that group in both scenarios.

Scenario	Average performance	Gate-keepers	Popular Scientists	Loyal Scientists	Embedded Scientists
Star Scientists Included	**3436.35**	**8.45**	**5.67**	**6.77**	**3.89**
Star Scientists Excluded	2951.62	4.31	2.76	4.97	2.56

Table 31: The performance per group with and without star scientists in the network

The structure of the Canadian nanotechnology network has been also analyzed in the scenario where star scientists were excluded from the network and the results were then compared to the original one. In both scenarios, we have calculated the average of degree centrality, betweenness centrality, and clustering coefficient (CC) as well as the network density.

Scenario	Average Betweenness	Average Degree	Average CC	Network Density
Star Scientists included	**0.0032**	**6.58**	**0.47**	**1.28**
Star Scientists excluded	0.0027	6.61	0.53	1.20

Table 32: The network structure in the scenarios of star scientists existence and absence

The results show no significant change in average degree centrality of the network for the two scenarios. In other words, the scientists will always seek for collaborative partnership even if their preferred partners are not there. As the density is related to the size of the network, the removal of some nodes (star scientist) would let to a smaller network size, the change in this measure for the both scenarios is relatively inconsiderable. That is, the proportion of ties in a network is comparable to the total potential ties, as shown in the table. Due to the centralized positions of the stars it was expected that their absence would decrease the degree centralization of the whole network. However, our results show that in both scenarios scientists will be engaged in collaborative activities with an equivalent opportunity to find partners (not necessarily star) scientists.

However, other network properties are slightly affected and changed when we excluded the star scientists. The average betweenness centrality, for example, reduces from 0.0032 to 0.0027, which means the overall centralization of the network will be negatively affected by removing the star scientists due to their centralized positions.

The change in network centralization will obviously affect the knowledge transmission among its nodes. The scenario, where star scientists are included, with higher average betweenness centrality has potentially more flows of knowledge between different network clusters. Consequently, the cliquishness of the network, represented by the average clustering coefficient, increased when nodes with central positions were removed. The 47 percent likelihood for two individuals with a common collaborator to also have partnership together when star scientists are in the network has been increased to almost 53% in case of their absence.

Overall, as it can be observed from simulation experiment results, we can conclude that the star scientists play a critical role in enhancing the scientific productivity of

Canadian nontechnology network. Their importance demonstrated by their individual performance, as active partners with high average number of publication, which will improve the overall network productivity. In addition, the knowledge diffusion in the network is affected by their centralized positions. The consequences of their absence are mainly to increase the cliquishness in the network and enhance the chance that other scientists be selected as partners and involved in collaboration activities.

5.4.2 The Role of Gatekeepers

Gatekeepers have a significant impact not only on the success of the networks, but also on the improvement of the performance of individuals connected to them in the network. Their role as controllers of the connections and resources in the network can even affect the direction of the research (Heikkinen *et al.*, 2007).

Based on our definitions in the present thesis, gatekeepers stand in the shortest paths of many other researchers, thus they facilitate the communication and knowledge flows in the network. Accordingly, the scientists with no direct connection to gatekeepers in the network have lower chance to be involved in collaboration activates with others, which may affect the overall network performance and knowledge flow as well. In this section, we analyzed the network productivity and structure after removing the gatekeepers completely from our model. The results have been also compared to the corresponding ones gained from the basic scenario (where all groups are there including gatekeepers).

As for the network performance, the average productivity per scientist is depicted in the figure below, showing that the average individual productivity in the network in the absence of Gatekeepers reduces over time to almost one third of the one with their existence. Scientists in the first scenario, with gatekeepers present in the network, have a better average performance of 1.66 article/author versus 1.26 article/author in the second scenario where the gatekeepers have been excluded.

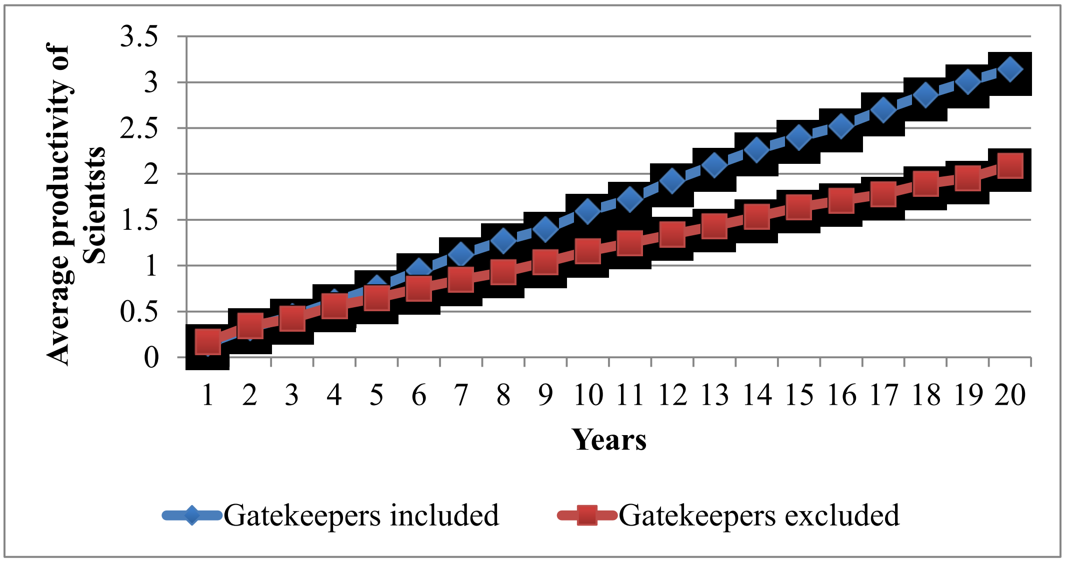

Figure 27: Average number of publications per scientist-Gatekeepers included/excluded

The overall network productivity is considerably affected by the existence of gatekeepers. As shown in the results of ten independent replications of the simulation model for two scenarios, the average of total number of publications per year is 3436.35 in a Gatekeeper-included network, while this number decreases to 2825.07 in a Gatekeeper-excluded network. The average numbers of publications in both scenarios for the ten replications are illustrated in the figure below. A negative impact of the gatekeepers' absence on the overall network productivity is shown in the results.

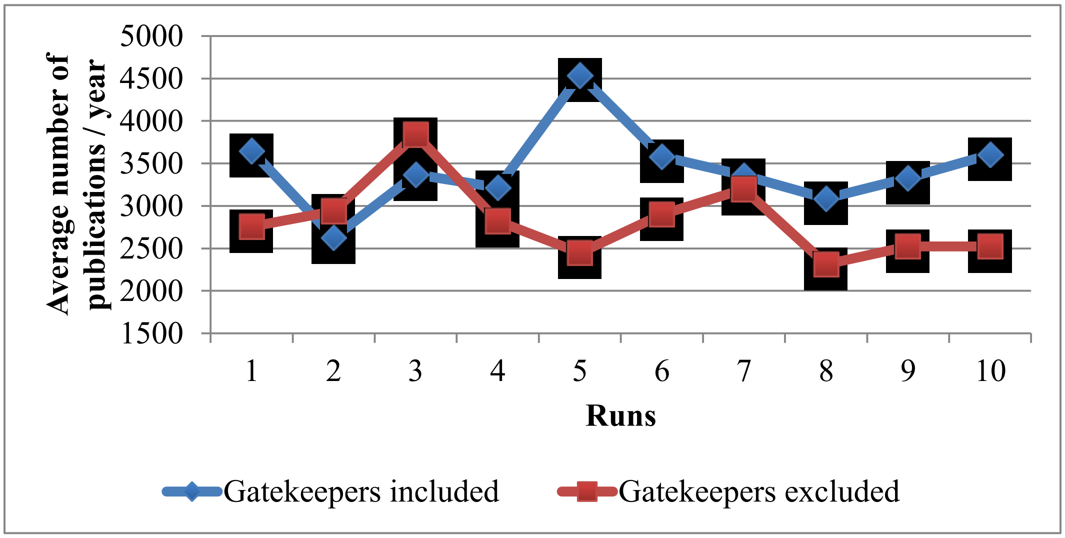

Figure 28:Average network productivity-Gatekeepers included/excluded

Similar to the stars, the less number of star scientists provides greater probability that they be partnered with more in more than one collaboration activity at the same time. That is, the fewer portions of gatekeepers in the world would enhance their individual performance and the overall productivity as well. The figure below shows the average individual productivity and the network performance in 5 different scenarios including the increase and decrease in the percentage of gatekeepers to the population. Both lower and higher ratio of gatekeepers to the population decreases the overall network productivity. Based on this result, along with the similar finding in the previous section about star scientists, we can suggest that the optimal percentage of the gatekeepers in the network to achieve the highest possible productivity is somewhere around 5%. We can also notice that increasing this percentage is even worse as it would lead to a wide distribution of knowledge among the network. We can say that when more scientists establish partnerships through the same gatekeepers result in enhancing the knowledge flows within the network by improving its centrality. As mentioned before, more research is needed to shed more light on the issue of the optimal number of these groups.

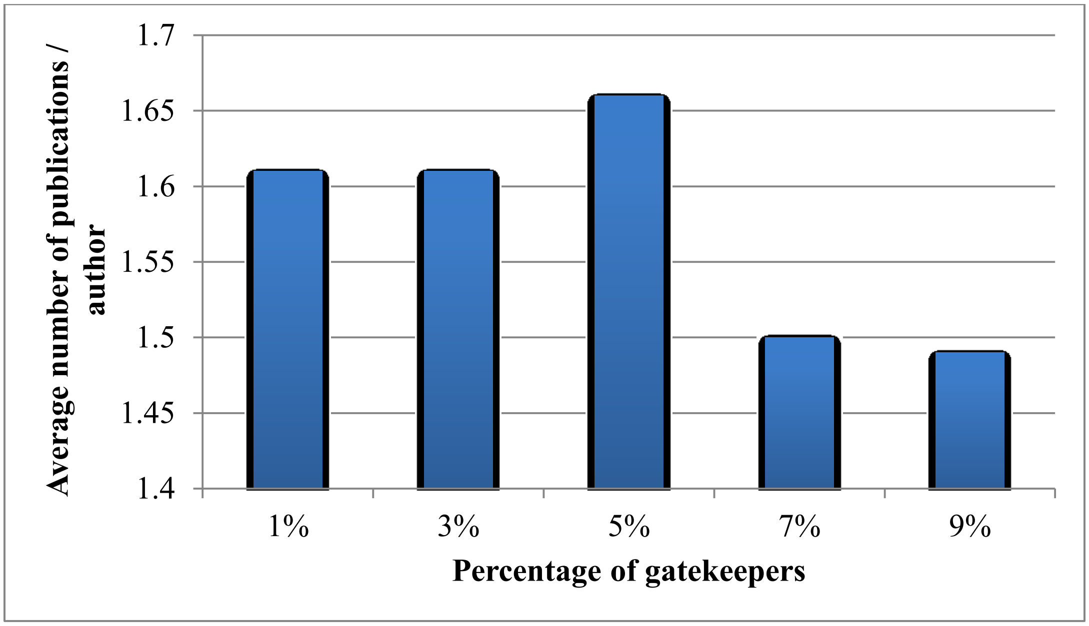

Figure 29: The average research performance with different percentage of gatekeepers

Each group's performance has been also analyzed to investigate how other scientists would behave in case of their absence. Due to their centralized position in the network, gatekeepers play critical role in bridging the knowledge transmission between scientists. That is, two scientists in different cluster (research group, institution or geographical area) who are directly connected to gatekeepers have a chance to exchange knowledge through the gatekeeper. In other words, gatekeepers are responsible for bring the knowledge into the research groups (network cliques) and are also more likely to be selected as a common partner between these groups.

Therefore, it is expected that the performance of other groups will change when we exclude gatekeepers from the network. The table below compares the average number of articles coauthored by scientists in each group of scientists in two scenarios, one including and one excluding the gatekeepers. As expected, we can see in the table a slightly negative impact on the performance of all groups except for the star scientists. When gatekeepers are not present in the network, other scholars still need to get the access to the information and turn to the most trustable and well-known ones within their cluster, i.e. star scientists. Star scientists are usually not only well connected but they are attractive partners in their own right. Star scientists thus seem to be playing the role of a substitute for the gatekeepers.

Scenario	Average performance	Star scientists	Popular Scientists	Loyal Scientists	Embedded Scientists
Gatekeepers Included	**3436.35**	**25.73**	**5.67**	**6.77**	**3.89**
Gatekeepers Excluded	2825.07	27.27	4.05	5.21	3.62

Table 33: The performance per group with and without gatekeepers in the network

Network properties have been calculated to analyze the structure of the gatekeepers-excluded network and then compared to the gatekeepers-included one. The changes in average of degree centrality, betweenness centrality, and clustering coefficient (CC) as well as the network density in the two scenarios have been evaluated.

Scenario	Average Betweenness	Average Degree	Average CC	Network Density
Gatekeepers Included	**0.0032**	**6.58**	**0.47**	**1.28**
Gatekeepers Excluded	0.0020	6.61	0.67	1.17

Table 34: The network structure in the scenarios of gatekeepers' existence and absence

Based on the definition of gatekeepers in the present thesis, i.e. the scientists with highest betweenness centrality, we were expecting a sharp reduction in the average network betweenness in case of their absence. The results confirm this hypothesis as the average betweenness has considerably decreased from 0.0032 to 0.0020 when gatekeepers have been removed from the network.

Gatekeepers in fact connect some clusters/cliques together. If they are not present in the network some of the researchers within the clusters will go to star scientists to seek the knowledge (which we suggested previously), but the others will probably start collaborating more within their own groups (clusters). Thus, since gatekeepers act as connection points in the network by having shortest paths running through them, their absence appears to result in higher network cliquishness. This is expressed by the considerably higher average clustering coefficient of 0.67 for the network without gatekeepers comparing to only 0.47 with gatekeepers there.

The presence of gatekeepers in the network results in lower number of ties between scientists required to generate the best possible knowledge transmission. Therefore, in the scenario where gatekeepers were excluded, each scientist will still try to gain access to the external knowledge by building up his/her own ties. Thus, there will be no considerable change in the average degree centrality in both scenarios. However, the network density, the proportion of ties in a network relative to the total potential ties, has decreased from 1.28 in the first scenario to 1.17 when gatekeepers were excluded. That is, when gatekeepers who provide the shortest paths are not present there will be more potential ties to accomplish network connectivity and knowledge diffusion.

Overall, the results of the simulation study suggest that the productivity and structure of the network is greatly impacted by its gatekeepers. As they act as bridges facilitating the knowledge exchange between clusters, their absence will result in slower flow of knowledge between nodes of the network and encourage the forming of closed research groups.

5.4.3 The Role of Popular Scientists

The scientists with a high number of connections in the network, i.e. with highest degree centrality, act as popular scientists in this study. Our interest in studying their role comes from the large number of collaborators they know, which obviously affects the knowledge sharing and transmission within the network. We want to recognize the behavior of the whole network and analyze its structure if scientists with high number of connections quit the network. For this purpose, we run a new scenario of our model using the appropriate setting that match this objective. We have excluded popular scientists and performed the analysis of the productivity and structure of the network, and compared the results to the ones of the world.

The average productivity per scientist, as an indicator for the network performance, is illustrated in the figure below, for both scenarios. The results show inconsiderable difference in the average individual productivity in the absence of popular scientists. The average performance in popular scientists-included network is 1.66 article/author where in the second scenario, i.e. when they were excluded; scientists have an average of 1.48 article/author.

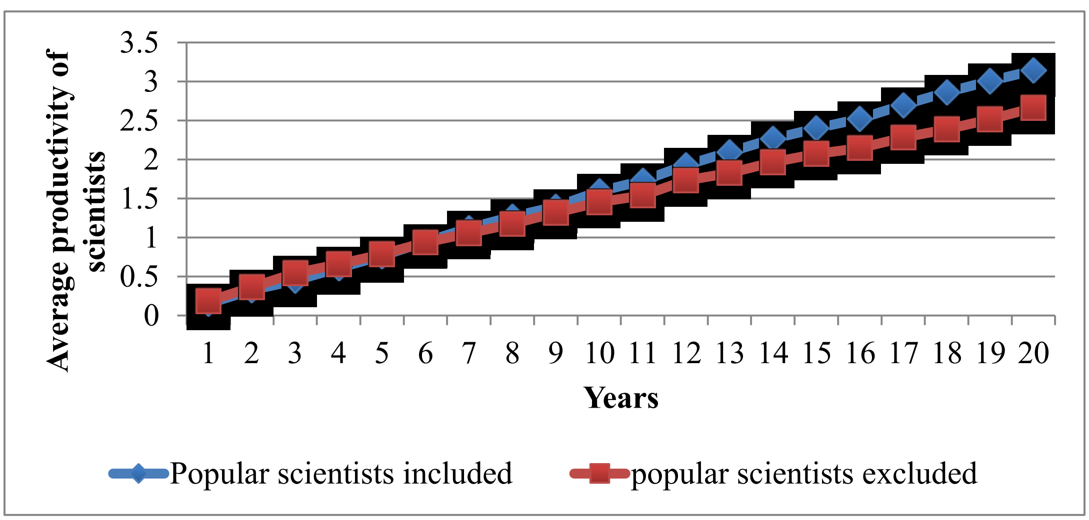

Figure 30: Average number of publications per scientist-Popular scientists included/ excluded

Although the absence of popular scientists does not have much impact on the individual performance, the overall performance has been negatively affected by their absence. The figure below shows the results of ten independent replications of the simulation model. The average of total number of publications per year is 3436.35 in the scenario where popular scientists included, has considerably reduced to 3051.34 publications/year when popular scientists have been removed from the network.

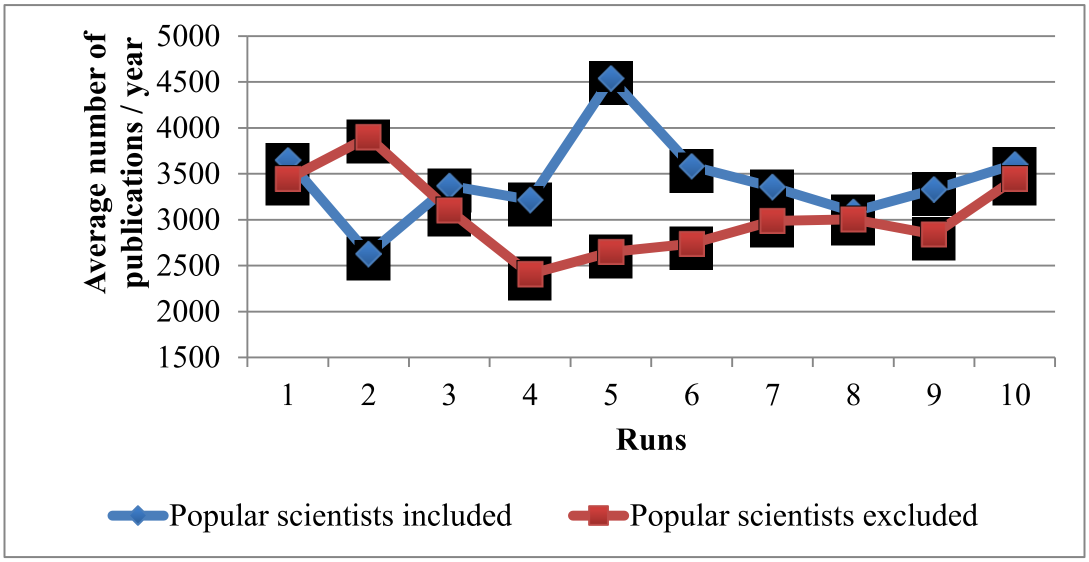

Figure 31: Average network productivity-Poplar scientists included/excluded

Although scientists keep almost the same level of performance, each collaboration activity will have less number of partners. In case of popular scientists absence the average number of scholars involved in a scientific partnership is 4.74 coauthors instead of 6.22 coauthors when they are there. However, having different percentage of popular scientists to the population would have inconsiderable difference in the average number of partners involved in each collaborative activity.

Our experiments showed an increasing efficiency of the network with more popular scientists. The results of the scenarios where we changed the ratio of this group to the population present a considerable improvement in the research performance in correlation with the increase of popular scientists. That is, the most well connected scientists we have in the network the more cohesive the network becomes and the knowledge follow is enhanced. The figure below shows the improvement in both individual and network productivity by the increased percentage of popular scientists.

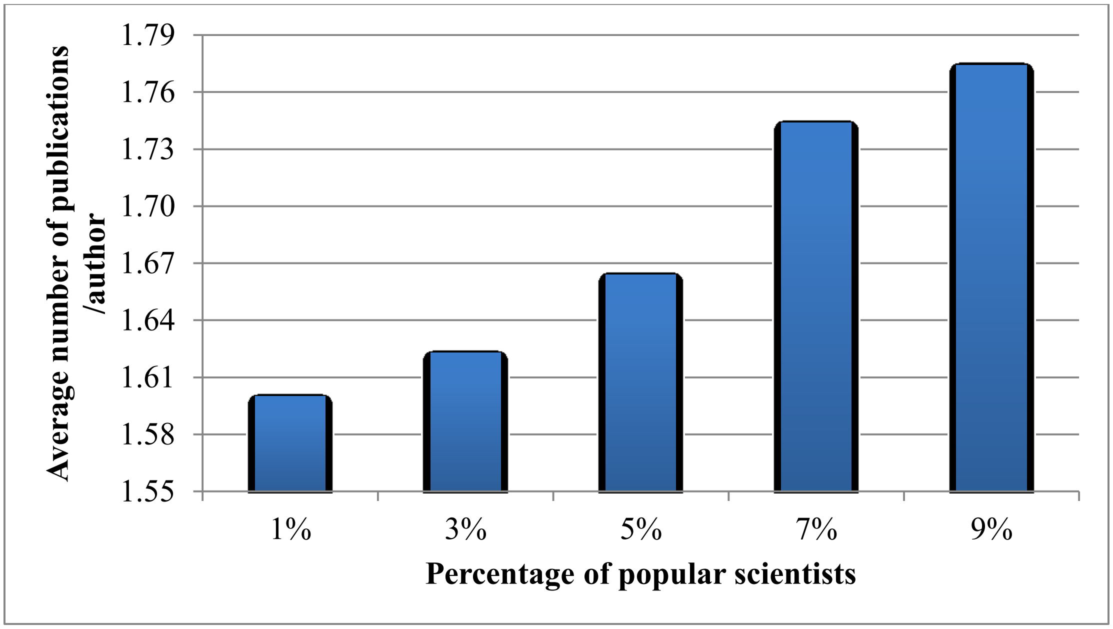

Figure 32: The average research performance with different percentage of popular scientists

Besides, it is expected that an overlapping exists between this group and others. That is, popular scientists with high number of connections can be also a part of star scientists and gatekeepers group who are most likely be selected as partners by all scientists. Consequently, the performance of these groups might also be affected by removing the popular scientists from the network. The table below shows the performance of each group of scientists measured by the average number of articles coauthored by them in both scenarios.

Scenario	Average performance	Star scientists	Gatekeepers	Loyal Scientists	Embedded Scientists
Popular Scientists Included	**3436.35**	**25.73**	**8.45**	**6.77**	**3.89**
Popular Scientists Excluded	3051.34	28.89	8.00	6.96	3.76

Table 35: The performance per group with and without popular scientists in the network

We have also studied the structure of the Canadian nanotechnology network calculating several network properties such as the average degree centrality, betweenness centrality, and clustering coefficient (CC) as well as the network density in the two scenarios. The table below presents these measurements of the popular scientists-excluded network and compares them to the ones measured in the network with popular scientists included.

Scenario	Average Betweenness	Average Degree	Average CC	Network Density
Popular Scientists Included	**0.0032**	**6.58**	**0.47**	**1.28**
Popular Scientists Excluded	0.0208	5.53	0.51	1.24

Table 36: The network structure in the scenarios of popular scientists existence and absence

As expected the average degree centrality of the network, as an indicator of the average number of collaborators per node, has decreased by from 6.58 to 5.53 for the second scenario. The fact that the portion of authors with extremely high degree centrality is very small comparing to our large population makes this relatively significant change even more considerable. Apparently, the flow of knowledge in the network is very much affected by sharing the knowledge between the nodes through collaboration activities.

The average betweenness centrality has been also negatively affected by the absence of popular scientists. That is, the average betweenness of the network that includes popular scientists is 0.0032, but this decreases to 0.0028 after their removal. Mathematically, a network with both higher average degree centrality and higher average betweenness centrality is more centralized and theoretically supports better flow of knowledge. Thus, we can conclude that by including the popular scientists in the network, the overall degree and betweenness centralization of the network improves, which subsequently enhances the knowledge flow within the network.

There are also slight changes in the other properties for the second scenario. The average clustering coefficient has increased when popular scientists were removed from the network indicating higher cliquishness. High network clustering affects the network flow negatively; as more clustered groups have many redundant links bearing the same knowledge and little fresh knowledge flowing to the cluster. The network density however is almost similar in both scenarios representing inconsiderable change in the portion of actual links comparing to the potential ones in the case of popular scientists absence.

Overall, the results of the simulation study suggest that although the overall productivity is not much affected by the absence of popular scientists, there is a significant impact on the structure of the network. Popular scientists have more connections and are thus greatly responsible for the knowledge sharing and involvement of the nodes within the network. However, in the network that does not have nodes with extremely high number of connections, the scientists continue performing with less number of partners involved in each collaboration activity. On the other hand, this group noticeably affects the centrality of the network. The absence of popular scientists will decrease the efficiency of the knowledge flow in the network and increase its clustering.

5.4.4 The Role of Loyal Scientists

Over 56% of the respondents to our survey stated that the prior satisfactory collaboration experience is a critical factor to be considered for partner selection mechanism. Accordingly, we are assuming that most of the researchers in our database, are actually considering this factor while selecting their partners. On the other hand, it has been discussed in the literature that loyalty, i.e. maintaining strong collaboration ties, causes the structure of collaborative networks to become embedded (Mat *et al.*, 2009). That means the flow of knowledge is expected to improve when loyal scientists, who are repeating collaborations with the same partners, are removed from the network. As discussed earlier in the present thesis, we have indicated the loyalty of scientists in our network by their weighted degree centrality, i.e. by the ratio of the sum of a node's link weights (number of co-authorships) to the total number of different collaborators. We have run the scenario where all scientists with the highest weighted degree are excluded from the network in order to study the impact of this setting on the productivity and structure of the network. Although seeking among the previous partners is a default collaboration strategy that is used in all scenarios, only the scientists with the highest collaboration frequency will be considered as loyal. In each step, the model will find those who became more loyal after the latest collaboration activities, and remove them. The figure below shows the individual performance comparing to the loyal-scientists included network. The results confirm the high research performance achieved by scholars with a strong relationship (frequent co-authorship). In the case of loyal scientists present, the average performance is 1.66

article/author where in the second scenario when they were excluded scientists have an average of 1.37 article/author.

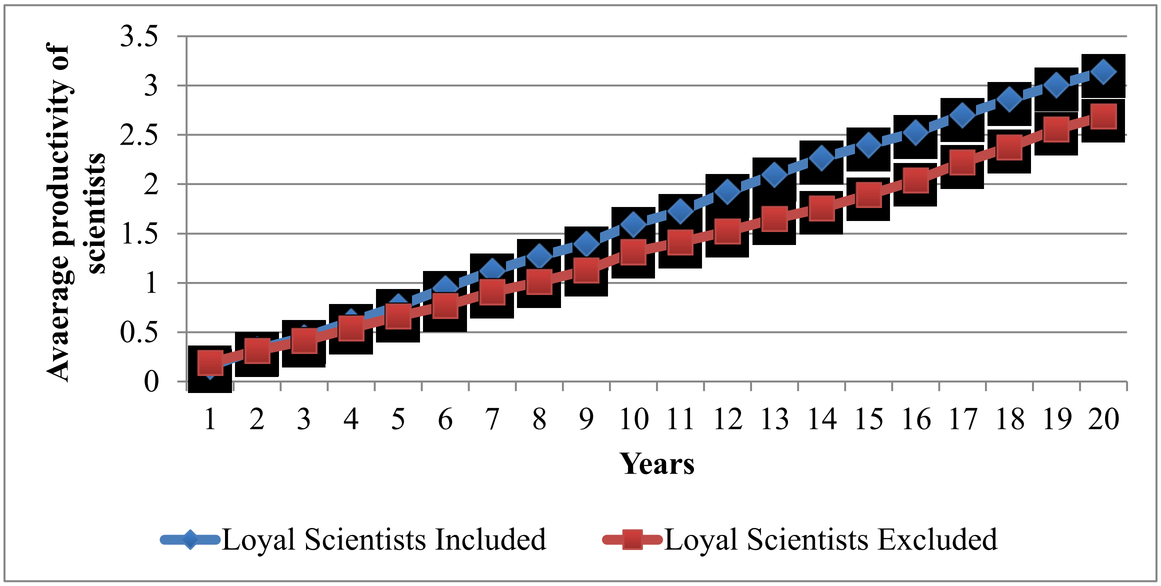

Figure 33: Average number of publications per scientist-Loyal scientists included/ excluded

The performance of the network is positively associated with the individual performance of scientists in the network. The figure below shows that the average of total number of publications per year is greatly affected by the frequency of repeated collaboration. An average of 3436.35 publications/year in the scenario where loyal scientists included has considerably reduced to 2676.13 publications/year when they excluded.

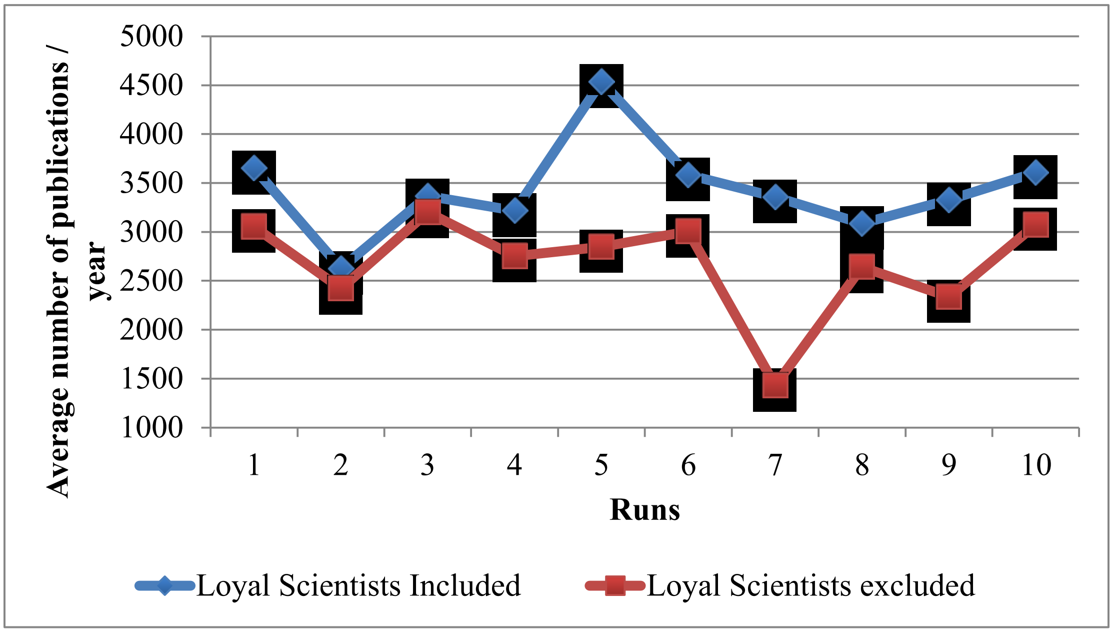

Figure 34: Average network productivity-Loyal scientists included/ excluded

The figure below shows the average productivity per scholar and the network performance in 5 different scenarios, which involve various percentages of loyal scientists to the population. A better performance of the network has been observed every time the number of the loyal scientists in the network has increased. In fact, the increasing number of scientists who have already satisfactory collaboration experiences would motivate them to renew the partnership and involve in new research activities together. This consequently affects the individual productivity as well as the overall network efficiency.

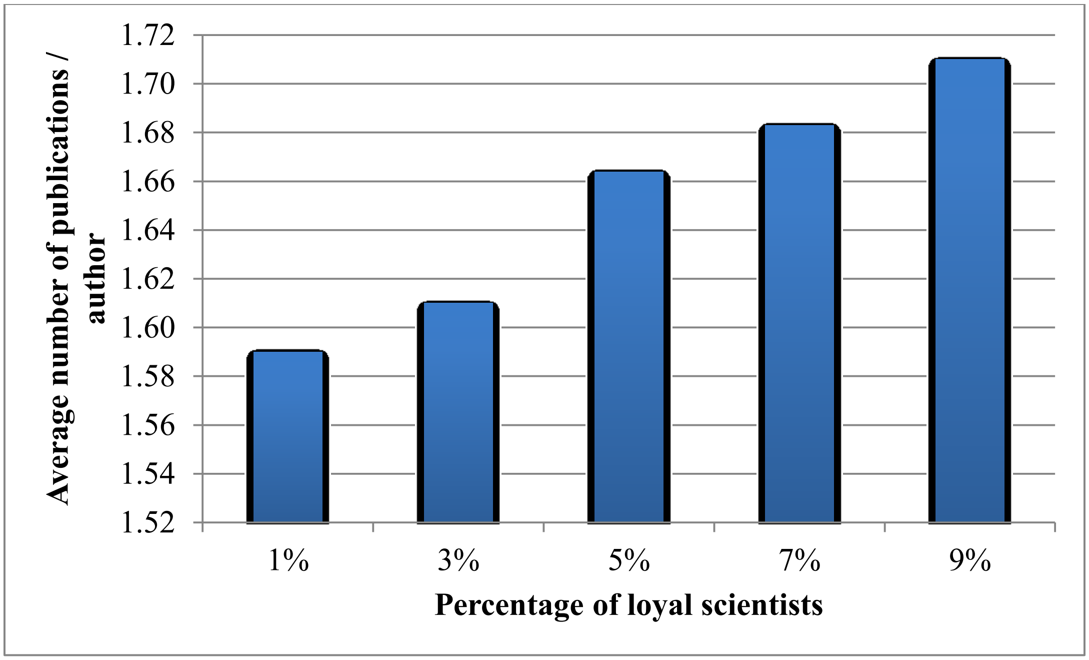

Figure 35: The average research performance with different percentage of loyal scientists

The performance of all the other groups might also be affected by removing the loyal scientists from the network. The table below compares the average number of articles coauthored by each group of scientists both in loyal scientist included and excluded scenarios. For most of the groups there was no significant change, except star scientists whose performance improved significantly after the loyal scientists were removed. Since we assume that the loyal scientists are among the productive researchers, their removal from the network will leave their former collaboration partners in need for some active productive researchers to collaborate with. The star scientists will thus again play a role of substitutes and create thus many new fruitful ties, which is reflected by their great jump in the performance.

Scenario	Average performance	Star Scientists	Gatekeepers	Popular Scientists	Embedded Scientists
Loyal Scientists Included	**3436.35**	**25.73**	**8.45**	**5.67**	**3.89**
Loyal Scientists Excluded	2676.13	33.01	8.46	5.61	3.90

Table 37: The performance per group with and without loyal scientists in the network

The structure of the Canadian nanotechnology network has been also mathematically analyzed in the two scenarios with and without loyal scientists. The table below represents average of degree centrality, betweenness centrality, and clustering coefficient (CC) as well as the network density have been calculated in both scenarios.

Scenario	Average Betweenness	Average Degree	Average CC	Network Density
Loyal Scientists Included	**0.0032**	**6.58**	**0.47**	**1.28**
Loyal Scientists Excluded	0.0029	6.66	0.40	1.26

Table 38: The network structure in the scenarios of loyal scientists existence and absence

Although there is no significant change in the betweenness centrality and density of the network in the absence of loyal scientists, the table above denotes that their absence would result in higher average degree centrality, and also lower average clustering coefficient in the network. The average degree centrality of 6.66 reflects higher number of connections for each node in case of excluding the loyal scientists comparing to 6.58 in case of their presence. In a network without loyal scientists there is higher possibility of generating more innovative ideas as a result of accessing to new knowledge through having partnership with new collaborators.

The table also shows that the clustering coefficient of the network in the absence of the loyal scientists is lower, indicating that loyal scientists get involved in strong collaboration relationship ties which makes the collaboration pattern more embedded and the network more cliquish. In a loyal scientists-included network, there is 47 percent chance for two individuals with a common collaborator to also have partnership together, whereas this chance is around 40 percent in the scenario when they have been excluded. This means that the collaborators of the loyal scientists start searching outside their close circles when their usual collaborators are not available. There is much higher probability that the knowledge they gain outside their clusters is new and fresh to him/her and to his/her collaborators, as opposed to the often-redundant information they gain within their collaboration clusters.

The increasingly loyal behavior within the network will result in more strong collaboration ties, and thus even more loyal scientists would appear. It is expected that with higher number of loyal scientists the impact on the knowledge flow will be getting more negative. The table below shows that the more loyal scientists comparing to the population will result in higher average clustering coefficient and lower average betweenness centrality for the whole network, both properties, which affect the network, flow negatively.

Loyal Scientists percentage	1%	3%	5%	7%	9%
Avg. Betweenness	0.0035	0.0033	**0.0032**	0.0030	0.0028
Avg. CC	0.40	0.42	**0.47**	0.55	0.60

Table 39: The average network properties with different percentage of loyal scientists

Eventually, based on the simulation results for the above-mentioned scenarios, we can conclude that although having strong collaboration ties within the network would enhance the individual research performance, its negative impact on the flow of knowledge is considerable. The results suggest that maintaining the collaboration relationship with the same partners worsens the knowledge transmission performance of the whole network by making the network less central and more embedded.

5.4.5 The Role of Embedded Scientists

In the present thesis, we have defined embedded scientists as the ones with highest probability that their directly connected partners will be also connected. That is mathematically represented by the highest clustering coefficient, which tells how much of a node's collaborators are, on average, willing to collaborate with each other. The embedded scientists assist their connections to involve deeply in a local network of collaboration (his/her research group). To study the importance of this group of scientists to the research performance and knowledge flow and transmission, we have studied two different scenarios. Scientists with the highest clustering coefficient have been completely removed from the network, and its productivity and structure have been compared to the corresponding results from the original scenario.

The productivity of the Canadian nanotechnology network has been measured by the average number of publications per scientists as shown in the figure below. The results indicate a slight improvement in the research performance of scientists in a network without embedded scientists over the time. The average performance in the original scenario, where all groups are included, is 1.66 article/author where it has increased to 1.77 article/author in the case of embedded scientists absence.

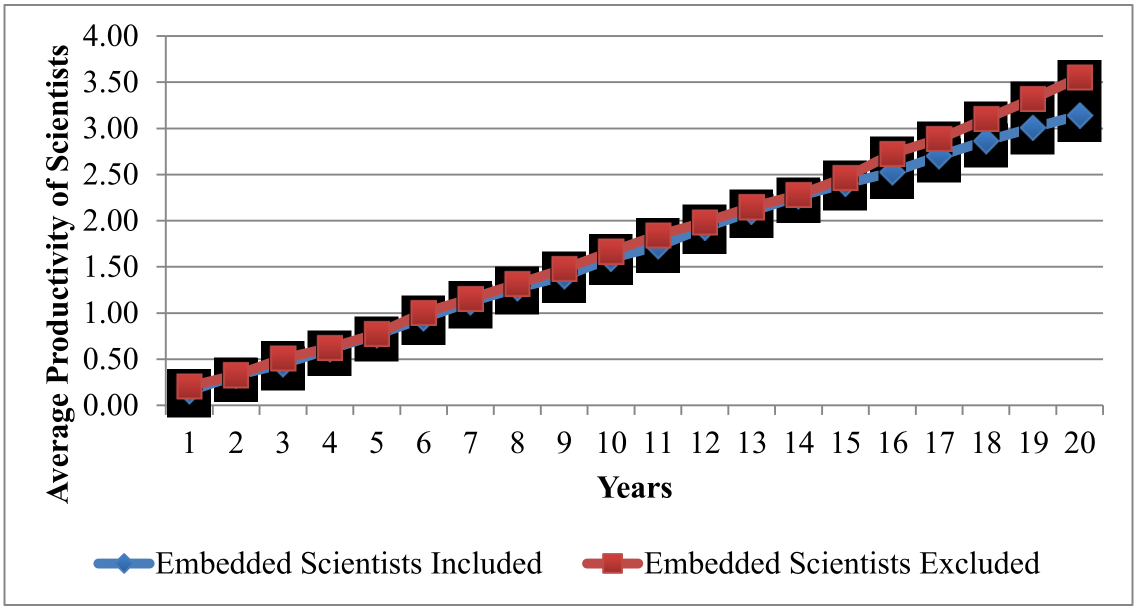

Figure 36: Average number of publications per scientist-Embedded scientists included/ excluded

Likewise, the overall performance of the network would slightly improve when embedded scientists excluded. The figure below shows that the average of total number of publications per year in both scenarios. An average of 3436.35 publications/year in the scenario where loyal scientists were included has considerably risen to 3524.73 publications/year when they were excluded.

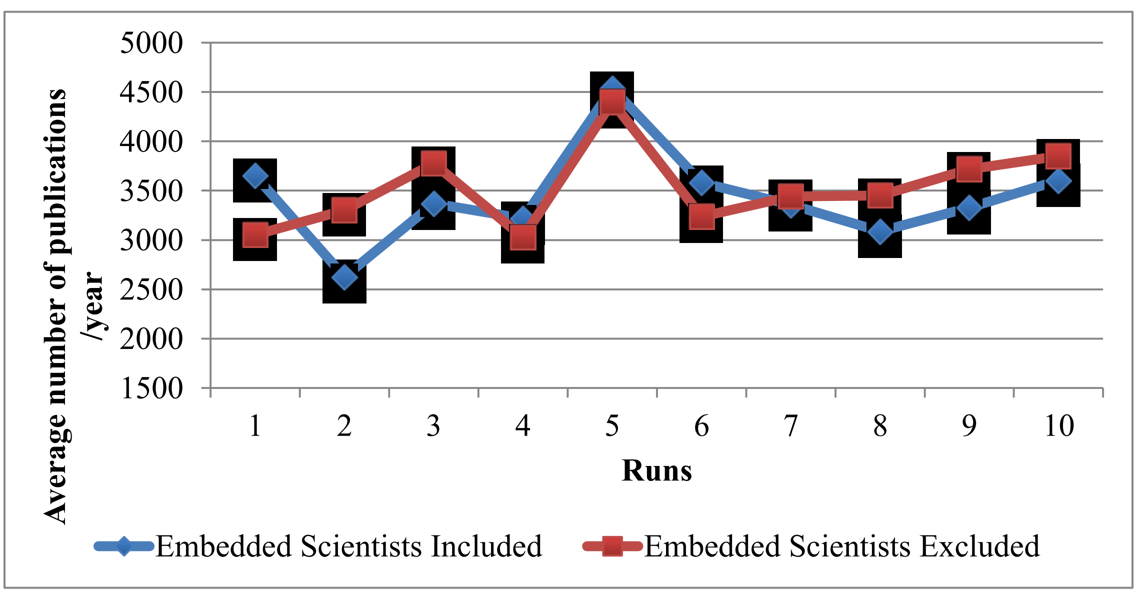

Figure 37: Average network productivity-Embedded scientists included/ excluded

As the efficiency of the network correlates inversely with its cliquishness, the increasing number of embedded scientists led to the presence of more clusters (closed research groups) and expected to negatively affect the knowledge transmission among the scientists. The figure below shows the result of different scenarios including lower and higher values for the percentage of embedded scientists comparing to the default setting of 5%. The observed finding supports the hypothesis that the lower number of embedded scientists within the scientists' population results in a better performance of the network, and thus increase the average productivity.

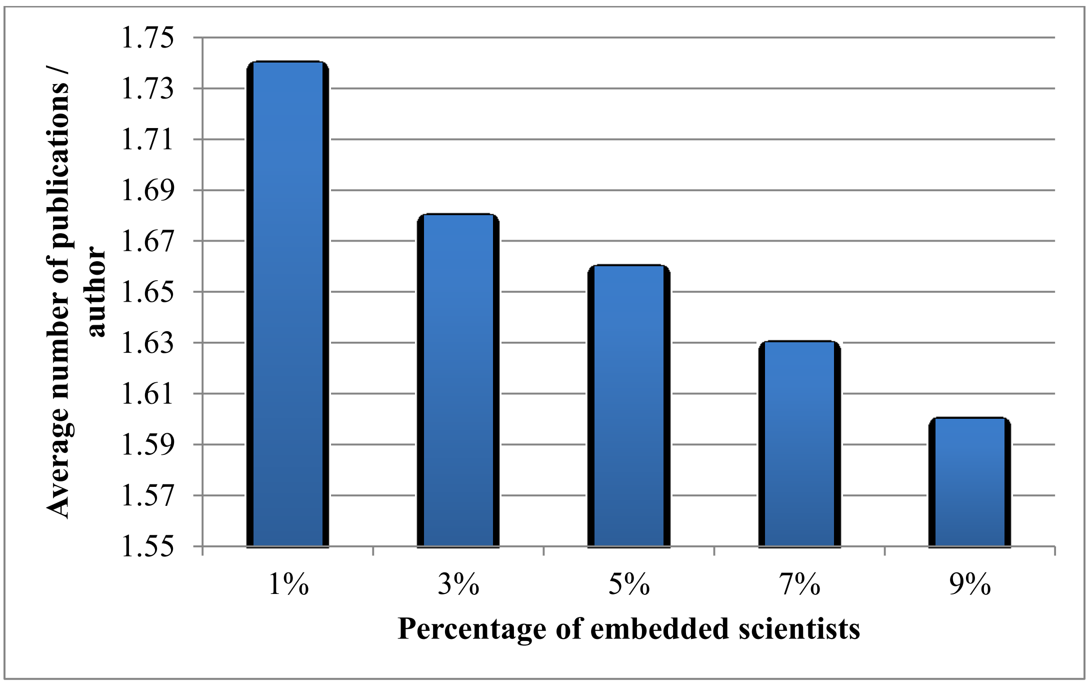

Figure 38: The average research performance with different percentage of embedded scientists

We have also studied the performance of other groups to see whether excluding embedded scientists from the network has an impact on others. The table below compares the average number of articles coauthored by scientists in each group of scientists in two scenarios, one with the embedded scientists' presence and one with their absence. We can see in the table a positive impact of their removal on the performance of all groups except for the gatekeepers where no change was observed. Excluding these scientists from the network thus opens new opportunities for the scientists to collaborate with new partners outside their research group, which leads to a better individual performance.

Scenario	Average performance	Star scientists	Gate-keepers	Popular Scientists	Loyal Scientists
Embedded Scientists Included	**3436.35**	**25.73**	**8.45**	**5.67**	**3.89**
Embedded Scientists Excluded	3524.73	33.07	8.45	6.03	7.75

Table 40: The performance per group with and without embedded scientists in the network

The structure of the Canadian nanotechnology network has also studied by calculating several network properties such as the average of degree centrality, betweenness centrality, and clustering coefficient (CC) as well as the network density in the two scenarios. The table below represents these measurements of the embedded scientists-excluded network and comparing to the one with embedded scientists included.

Scenario	Average Betweenness	Average Degree	Average CC	Network Density
Embedded Scientists Included	**0.0032**	**6.58**	**0.47**	**1.28**
Embedded Scientists Excluded	0.0040	6.56	0.35	1.31

Table 41: The network structure in the scenarios of embedded scientists existence and absence

The table above indicates slight changes in the average degree centrality and density of the network in the absence of embedded scientists, which are inconsiderable. However, as expected, their absence would result in lower cliquishness and improved knowledge transmission performance of the whole network. Since the embedded scientists are identified in the network by a high clustering coefficient, their removal from the network must obviously cause the network to become less clustered. The lower average clustering coefficient of 0.35 suggests a lower probability of two individuals with a common collaborator to also have partnership together comparing the first scenario whereas this probability is 0.47. Consequently, the researchers will have more chances to gain external knowledge instead of being limited within a closed research group. A positive impact of their exclusion on the network structure can be also observed through the increase in the average betweenness centrality to 0.0040 versus 0.0032 for the network that includes the embedded scientists.

As mentioned before, the embedded scientists encourage the formation of the cliques by providing a chance for those who are directly connected to them to collaborate within their closed group. The more nodes in the network with higher probability to have their neighbors also connected, the less efficient knowledge flow can be expected. The table below shows average clustering coefficient, indicating the cliquishness, and the average betweenness centrality, indicating the network centralization, in 4 different scenarios with changing the default setting to more and less embedded scientists to exist in the network. The results support the previous finding that the more embedded sciences are present in the network the worse the network structure becomes, in terms of increasing cliquishness and decreasing betweenness.

Embedded Scientists percentage	1%	3%	5%	7%	9%
Avg. Betweenness	0.036	0.033	**0.032**	0.030	0.020
Avg. CC	0.40	0.44	**0.47**	0.53	0.67

Table 42: The average network properties with different percentage of embedded scientists

We can therefore conclude, based on the simulation results, that the presence of embedded scientists has negative impact on the average productivity per scientist in the network. Excluding these scientists from the network results in a better individual performance by opening new opportunities for the scientists to collaborate with new partners outside their research group. Some of the network characteristics are comparable for the two scenarios. However, in a network without embedded scientists a better flow of knowledge amongst scientists resulting in an enhanced growth of innovativeness was observed. By excluding the embedded scientists from the network, it becomes more centralized in terms of its betweenness and less clustered, which will support the knowledge exchange among clusters and reduces the number of closed research groups within the network.

6.0 Summary and Discussion

This chapter provides a review and discussion of the main findings of this study. It devoted to summarize the results of our simulation experiments and to a compare the variables of interest in all different scenarios. Our findings can be categorize into three main groups; the overall scientific production performance, the productivity of each individual group of scientists and the network structure.

As for the overall efficiency of the network we consider two indicators of performance of scientists (measured by the average number of publications/author) as well as the network productivity (measured by the average number of publications/year). The observed decrease in the network productivity and individual performance caused by the absence of star scientists, gatekeepers, popular scientists and loyal scientists suggest their critical contribution in enhancing the scientific production. Whereas the absence, or a lower number of, embedded scientists showed a better performance for both individual and network level.

It is interesting that although the complete removal of star scientists from the network resulted in a poor performance, including a higher percentage of them is not as good either. We can suggest that some optimum, which is not that high, may probably be there. Similarly, when we increase the portion of gatekeepers to the population, it does not always increase the performance, so some optimum percentage is again observable. Here we suggest that it may be around 5%, but more research would be needed to verify this.

Moreover, the increase of connectivity and loyal behavior in the network, represented by both popular and loyal scientists respectively, enhances the performance of the scientific production within the knowledge-based networks. The table below summarizes the values of the studied performance indicators resulting from the simulated scenarios and suggests the role of each group in enhancing the overall network efficiency.

Scenario	Publications/Scientist	Publications/Year
All groups included	**3436.35**	**1.66**
Star scientists excluded	2951.62	1.26
Gatekeepers excluded	2825.07	1.26
Popular Scientists excluded	3051.34	1.48
Loyal Scientists excluded	2676.13	1.37
Embedded Scientists excluded	3524.73	1.77

Table 43: Summary results for the overall network performance in different scenarios

The figure below illustrates the impact of a group's absence on the performance of others. The productivity of each group has been examined under all the tested scenarios; (scenario 1) where all groups are included in 5% each, (scenario 2) star scientists excluded, (scenario 3) gatekeepers excluded, (scenario 4) popular scientists excluded, (scenario 5) loyal scientists excluded, and (scenario 6) embedded scientists excluded.

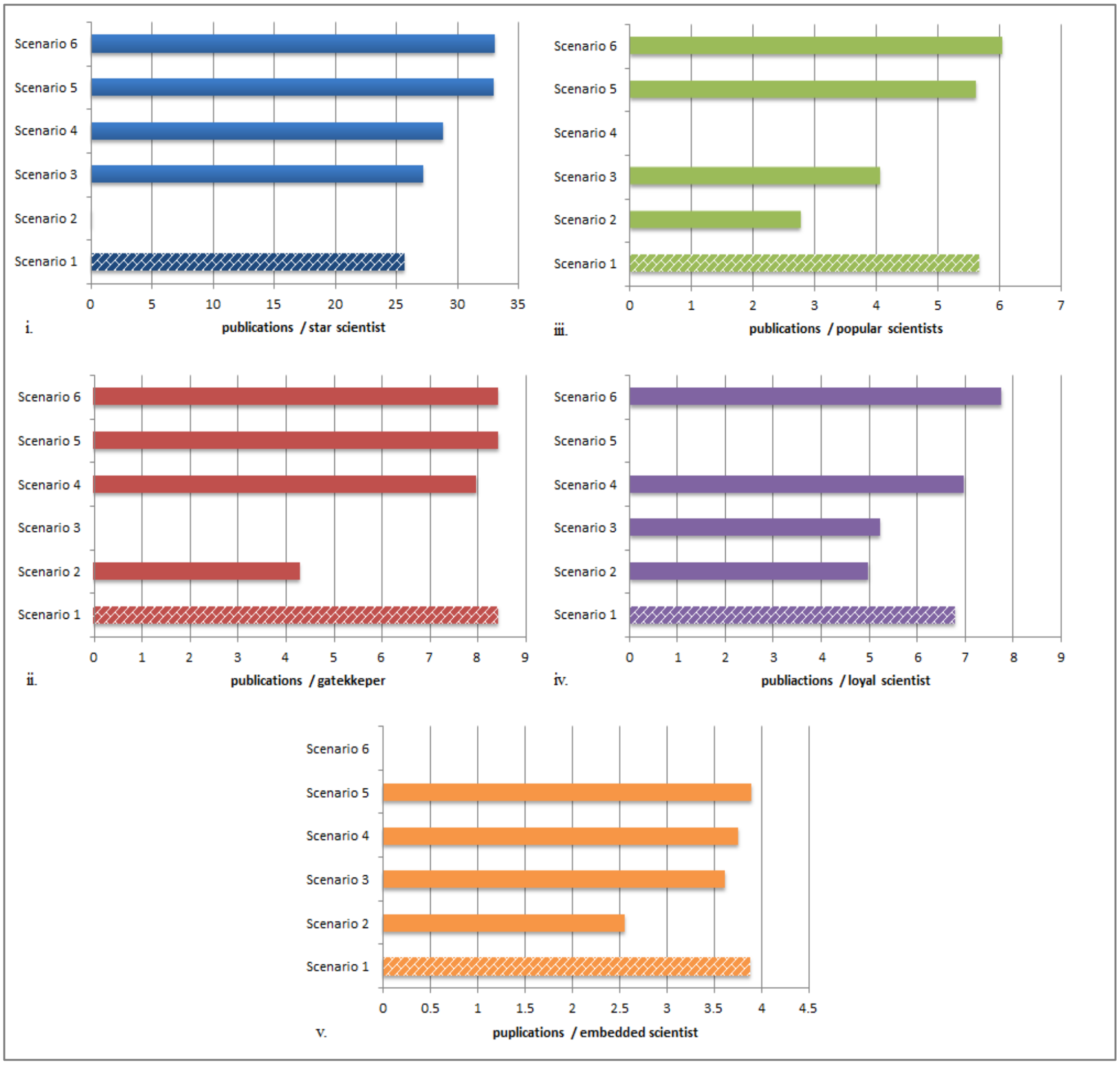

Figure 39: The performance of each individual group under different tested scenarios; (i.) star scientists, (ii.) gatekeepers, (iii.) popular scientists, (iv.) loyal scientists, (v.) embedded scientists.

The results suggest that scientists in each group are behaving differently under each scenario. For instance, a negative impact of the performance of all groups is observable when the star scientists, gatekeepers or popular scientists are not there. However, star scientists appear to play a substitutive role in the network, i.e. they are the ones most likely to be selected as potential partners, if the usual collaborators are missing. This role leads to an increasing productivity of the star scientists group in case that any other group is excluded.

Although the absence of loyal scientists results in no observable effect on the performance of others, the average productivity of almost each group is enhanced when the embedded scientists group is missing. The table below summarizes the values of average number of publications per group under the 6 tested scenarios.

Scenario	Star scientists performance	Gatekeepers performance	Popular Scientists performance	Loyal Scientists performance	Embedded Scientists performance
All groups included	**25.73**	**8.45**	**5.67**	**6.77**	**3.89**
Star scientists excluded	-	4.31	2.76	4.97	2.56
Gatekeepers excluded	27.27	-	4.05	5.21	3.62
Popular Scientists excluded	28.89	8.00	-	6.96	3.76
Loyal Scientists excluded	33.01	8.46	5.61	-	3.90
Embedded Scientists excluded	33.07	8.45	6.03	7.75	-

Table 44: Summary results for the groups' performance in different scenarios

Similarly, in order to compare the variation in different network indicators resulting from the absence of each group we present the results in the figure below. The figure shows the values for several network properties describing the changes in the overall network structure under the tested scenarios mentioned above.

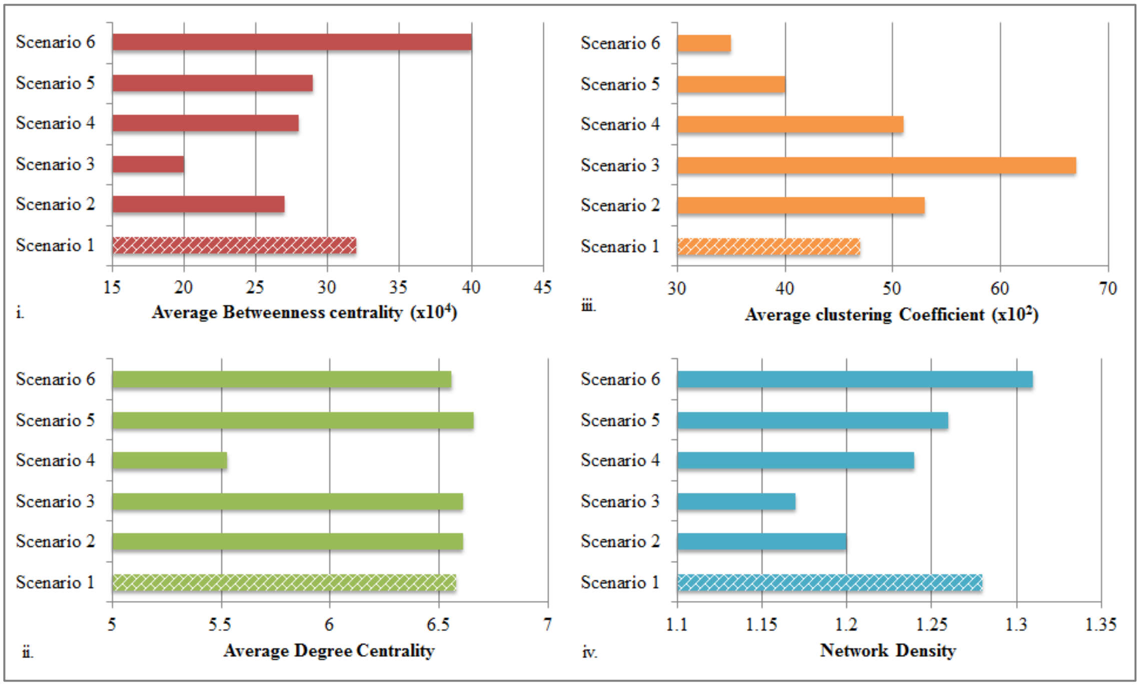

Figure 40: The network structure measurement under different tested scenarios; (i.) Network betweenness centrality x10^4, (ii.) Network degree centrality, (iii.) Network clustering coefficient x10^2, (iv.) Network density

Although a negative effect on the knowledge transmission though the network is encountered if loyal scientists and embedded scientists are present, the flow of knowledge is in fact positively affected by the presence of star scientists, gatekeepers and popular scientists. The connectivity of the network, i.e. average degree centrality, on the other hand is comparable for all scenarios except when popular scientists, the ones with highest number of connections, are missing.

Network density depends mainly on the size of the network, so there is no significant change in its value by removing nodes with specific network characteristics, except for the case of the gatekeepers' absence. As gatekeepers, nodes with highest betweenness, shape the shortest paths between nodes, their absence results in more possible links within the network, which results in lower density considering having the same number of actual links in both scenarios. The network structure properties under the 6 tested scenarios are summarized in the following table:

Scenario	Avg. Betweenness Centrality	Avg. Degree Centrality	Avg. Clustering Coefficient	Network Density
All groups included	**0.0032**	**6.58**	**0.47**	**1.28**
Star scientists excluded	0.0027	6.61	0.53	1.20
Gatekeepers excluded	0.0020	6.61	0.67	1.17
Popular Scientists excluded	0.0028	5.53	0.51	1.24
Loyal Scientists excluded	0.0029	6.66	0.40	1.26
Embedded Scientists excluded	0.0040	6.56	0.35	1.31

Table 45: Summary results for the network structure properties in different scenarios

7.0 Remarks and Recommendations for Future Work

7.1 Concluding Remarks

The main objective of this work is to evaluate the knowledge flows and transmission within the Canadian nanotechnology scientific production network. Our concern is to study the network at individual level to investigate the role of scientists and their collaborations in enhancing the innovative and research performance. This work has been done in two phases; first involves the analysis of the behavior of scientists in real world and the second one the simulation of the whole system under the control of various parameters.

The analyzed dataset has been extracted from SCOPUS database using specialized keywords related to nanotechnology where at least one of the co-authors was affiliated to a Canadian institution. We have created the knowledge-based network based on the co-authorship relationships between the scientists. This network has been visualized and mathematically analyzed using the social network analysis tool PAJEK.

The structural properties for each node (representing an author) in the network have been calculated and we studied the correlation between these properties and their research performance indicators including the number of publications, citations count and h-index. We have categorized the scientists into five groups based on their research performance and their positions in the network. The introduced groups are star scientists, gatekeepers, popular scientists, loyal scientists and embedded scientists. The highest values for number of publications, betweenness centrality, degree centrality, weighted degree centrality and clustering coefficient have been used as criteria to identify the scientists belonging to each group respectively.

A statistical and data mining analysis has been performed to detect a pattern for the research performance and collaboration behavior of each group. In terms of productivity, we have found that, as expected, star scientists (followed by gatekeepers) have the best performance where embedded scientists have the lowest. Regarding the collaboration behavior, star scientists and gatekeepers have the weakest collaboration ties, i.e. they have much single collaboration with many partners, which means that they are more likely to attract new scholars for scientific partnership.

In order to improve our understanding of the partners' selection mechanism, we ran a survey with an objective to elucidate the reasons behind the selection of potential collaborators. Questionnaire is sent to the previously identified active researchers in our database who have a scientific collaboration history. The findings show that the most critical factors to be considered while selecting the partners are: their academic reputation, their experience in a complementary field, the resources and funding accessibility, the previous collaboration relation with them and its strength. Moreover, we found that the personal factors such as gender, age, language and cultural background have no impact on the partners' selection decision.

To study the dynamics of the nanotechnology scientific production network, we have developed an agent-based model using NetLogo. We have applied the conceptual model that has been built in the first phase including as much knowledge as possible in our computerized model. Several techniques have been used to vivificate and validate the model and we have found that it has a sufficiently acceptable level of accuracy.

We have carried our various simulation scenarios to study the role of each identified group of scientists first by a complete removal of each group from the network and then by increasing and decreasing their ratio to the population. The results in each scenario have been analyzed concerning the impact of the changed settings on the research performance and structure of the network.

For the star scientists, we proved their critical role in enriching both the scientific production and knowledge flows of Canadian nontechnology network due to their high individual performance as well as their centralized positions. Stars also affect the knowledge diffusion in the network as active partners, who are attractive to be selected by other scientists, which will however reduce the chance for other scientists to establish partnerships.

The flow of knowledge within the network is highly affected by gatekeepers who are facilitating the knowledge exchange between various clusters of scientists in the network. They are responsible for bringing new knowledge into otherwise relatively closed research groups, by which they decrease the formation of the isolated network clusters in the network. Their outstanding individual performance contributes positively toward the improvement of overall productivity of the network.

The high numbers of connections that popular scientists have provide them with a unique role in increasing the speed of the knowledge sharing and transmission, enhancing connectivity within the network and decreasing its embeddedness. Consequently, it was surprising to find that the overall productivity of the network is not affected much by increasing or decreasing the number of partners involved in each collaboration activity.

Loyalty, i.e. maintaining strong collaboration ties, has a considerable negative impact on the flow of knowledge. Although the loyal scientists showed a good scientific production thereby greatly helping to improve the network productivity, the results suggest that maintaining the collaboration relationship with the same partners negatively affects the network structure over the time. That is, strongest collaboration ties would make the network more embedded and consequently worsen the knowledge transmission.

Embedded scientists provide higher chance for their collaborators to be involved deep in closed research groups. Thus, their individual performance would be lower than when they collaborate with new partners outside their research group. The results show the negative impacts of embedded scientists including making the network less centralized and more embedded. With lower average betweenness centrality, and higher average clustering coefficient the knowledge flow among clusters would be slow and the number of closed research groups within the network would be increased.

7.2 Limitations and Directions for Future Research

The contributions of this research were the essential first steps towards studying the performance of knowledge-based networks at the individual level. Many real-world problems were simplified or ignored due the need for more data or because their solutions were outside the scope of this research. In the following, few limitations of this study will be summarized and the opportunities for future research will be outlined accordingly.

First of all, although this work is mainly concerning the nontechnology industry in Canada, our developed model is sufficiently fixable to be used for extending the results of this research into the global level and/or comparing the findings to the comparable ones from other high-tech industries in Canada. On the other hand, further research could use more comprehensive database(s) where more information about the field of expertise, research interests and funding amount each scientist receives could be collected to improve the partner's selection mechanism in the model and reduce the level of randomness.

In addition, the analysis of network performance in our simulation model considers only quantity of the knowledge diffusion and transmission in nanotechnology field, i.e. average number of publications, and not the quality. Research performance indicators for individual scientists, such as the H-index, and for the research society, such as the RC-index and the CC-index, should be included for qualitative analysis.

Moreover, our results suggested that the absence of both star scientists and gatekeepers negatively affect the network performance the high percentage of them is not that good either. This could be an interesting issue for the scholarly to investigate if any optimum portion of this group should be there to achieve to possible efficiency for the network.

Furthermore, it would be interesting and more realistic to consider some details about the scientists' research career, i.e. change in their positions and/or mobility between different firms or organization. These changes might affect their productivity and open new opportunities for scientific partnerships.

Lastly, it is recommended that a more detailed study on partnership motivations would search into the reasons behind the observed different importance levels that various factors have as for scientists from different fields, affiliations and research experience.

References

[1.] Abbasi, A., & Altmann, J. (2011). On the correlation between research performance and social network analysis measures applied to research collaboration networks. *In 44th Hawaii International Conference on Systems Science (HICSS-44),* Jan. 4-7, Hawaii, USA.

[2.] Abbasi, A., Altmann, J., & Hwang, J. (2010). Evaluating scholars based on their academic collaboration activities: Two indices, the Rc-index and the Cc-index, for quantifying collaboration activities of researchers and scientific communities. *Scientometrics,* 83(1), 1–13.

[3.] Abbasi, A., Chung, K. S. K., & Hossain, L. (2012). Egocentric analysis of co-authorship network structure, position and performance. *Journal of Information Processing & Management*, 48(4), 671-679.

[4.] Ahrweiler, Gilbert, N. & Pyka, A. (2004). Simulating knowledge dynamics in innovation networks (SKIN) (No. 267).

[5.] Ahrweiler, P., Gilbert, N. & Pyka, A. (2011a). A New Model for University-Industry Links in Knowledge-Based Economies. *Journal of Product Innovation Management*, 28(2), 218-235.

[6.] Ahrweiler, P., Gilbert, N., & Pyka, A. (2011b). Agency and structure: a social simulation of knowledge-intensive industries. *Computational and Mathematical Organization Theory Journal*, 17(1), 59-76.

[7.] Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative science quarterly*, 45(3), 425-455.

[8.] Albino, V., Carbonara, N., & Giannoccaro, I. (2003). Coordination mechanisms based on cooperation and competition within industrial districts: An agent-based computational approach. *Journal of Artificial Societies and Social Simulation*, 6(4).

[9.] Albino, V., Carbonara, N., & Giannoccaro, I. (2006). Innovation in industrial districts: An agent-based simulation model. *International Journal of Production Economics*, 104(1), 30-45.

[10.] Ali, W., & Moulin, B. (2005). 2D-3D multiagent geosimulation with knowledge-based agents of customers' shopping behavior in a shopping mall. *In Spatial Information Theory* (pp. 445-458).

[11.] Alizadeh, D. and Schiffauerova, A. (2012) The Role of Individuals in Innovation Networks: A Simulation Approach in Canadian Biotechnology Network, *7èmes Journées de la Proximité*, Montreal, May 21-23, 2012

[12.] Alizadeh, Dorsa (2012) *The Role of Individuals in Innovation Networks: A Simulation Approach in Canadian Biotechnology Network.* Masters thesis, Concordia University.

[13.] Allen, R. (1983) Collective Invention. *Journal of Economic Behavior and Organization* 4, 1–24.
[14.] Alvarez, H. (2012). Collaborative Knowledge Creation. (May), 1-31.
[15.] Amaral, L. A. N., Scala, A., Barthélémy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97(21), 11149-11152.
[16.] Anklam, P. (2003). Tapping social networks to leverage knowledge and innovation.
[17.] Axelrod, R. (1997a). Advancing the art of simulation in the social sciences. *Research Policy*, 3(2), 16-22.
[18.] Axelrod, R. (1997b). The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration (*Princeton Univ. Press, Princeton, NJ*).
[19.] Balconi, M., Breschi, S., & Lissoni, F. (2004). Networks of inventors and the role of academia: an exploration of Italian patent data. *Research Policy*, 33(1), 127-145.
[20.] Bankes, S. C. (2002). Agent-based modeling: A revolution?. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3), 7199-7200.
[21.] Banks, J. (1998). *Handbook of simulation* (pp. 3-389). New York: Wiley.
[22.] Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *science*, 286(5439), 509-512.
[23.] Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3), 590-614.
[24.] Barabási, Albert-La´ szlo´ (2002). Linked: the New Science of Networks. *Cambridge, MA: Perseus Publishing*.
[25.] Baum, J. A., Calabrese, T., & Silverman, B. S. (2000). Don't go it alone: Alliance network composition and startups' performance in Canadian biotechnology. *Strategic Management Journal*, 21(3), 267-294.
[26.] Beaver, D. D. (2001). Reflections on Scientific Collaboration (and its study): Past, Present, and Future. *Scientometrics*, 52(3), 365-377.
[27.] Beaver, D., and Rosen, R. (1979). Studies in Scientific Collaboration Part III. Professionalization and the Natural History of Modern Scientific Co-authorship. *Scientometrics*, 1(3), 231-245.
[28.] Benner, M., & Sandström, U. (2000). Institutionalizing the triple helix: research funding and norms in the academic system. *Research Policy*, *29*(2), 291-301.
[29.] Berryman, M. J. & Angus, S. D. (2010). Tutorials on agent-based modelling with NetLogo and network analysis with Pajek. *In R. L. Dewar & F. Detering (Eds.), Complex Physical, Biophysical and Econophysical Systems: Proceedings*

of the 22nd Canberra International Physics Summer School (pp. 351-375). USA: World Scientific Publishing.

[30.] Birnholtz, J. P., & Bietz, M. J. (2003). Data at work: supporting sharing in science and engineering. In *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work* (pp. 339-348). ACM.

[31.] Birta, L. G., & Arbez, G. (2007). Modelling and simulation: exploring dynamic system behaviour. *Springer-Verlag New York Incorporated*.

[32.] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics reports*, 424(4), 175-308.

[33.] Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 3), 7280-7287.

[34.] Bosnjak, M., & Tuten, T. L. (2001). Classifying response behaviors in Web-based surveys. *Journal of Computer-Mediated Communication*, 6(3).

[35.] Bozeman, B., & Corley, E. (2004). Scientists' collaboration strategies: implications for scientific and technical human capital. *Research Policy*, 33(4), 599-616.

[36.] Breschi, S., & Lissoni, F. (2003). Mobility and social networks: Localised knowledge spillovers revisited. *Milan: University Bocconi, CESPRI Working Paper*, (142).

[37.] Breschi, S., & Lissoni, F. (2005). " Cross-Firm" Inventors and Social Networks: Localized Knowledge Spillovers Revisited. *Annales d'Economie et de Statistique*, 189-209.

[38.] Breschi, S., & Lissoni, F. (2006). Mobility of inventors and the geography of knowledge spillovers. New evidence on US data. *CESPRI WP*, (184).

[39.] Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., ... & Wiener, J. (2000). Graph structure in the web. Computer networks, 33(1), 309-320.

[40.] Bross, U., & Zenker, A. (1998). The Performance of Innovation Networks in Transition Economies : An Empirical Study of Slovenia, (September).

[41.] Buchan, N. R., Croson, R. T., & Dawes, R. M. (2002). Swift Neighbors and Persistent Strangers: A Cross- Cultural Investigation of Trust and Reciprocity in Social Exchange1. *American Journal of Sociology*, 108(1), 168-206.

[42.] Cantner, U., Conti, E., & Meder, A. (2010). Networks and innovation: the role of social assets in explaining firms' innovative capacity. *European Planning Studies*, 18(12), 1937-1956.

[43.] Carlsson, B., Jacobsson, S., Holmén, M., & Rickne, A. (2002). Innovation systems: analytical and methodological issues. *Research policy*, *31*(2), 233-245.

[44.] Chung, K. S. K., & Hossain, L. (2009). Measuring performance of knowledge-intensive workgroups through social networks. *Project Management Journal*, *40*(2), 34-58.

[45.] Cohen, J., & Linton, M. (2003). The Next Frontier for HIV/AIDS. Science, 301 (5640), 1650-Congressus numeratium, 158, 202-212.

[46.] Cowan, R., & Jonard, N. (2003). The Dynamics of Collective Invention. *Journal of Economic Behavior & Organization*, 52(4), 513–532.

[47.] Cowan, R., Jonard, N., & Zimmermann, J. (2007). Evolving networks of inventors. *Innovation, industrial dynamics and structural transformation*, 129-148.

[48.] Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web surveys: Perception of burden. *Social Science Computer Review*, 19, 146–162.

[49.] Currie, S.M. & Cheng, C.H. (2013). A practical introduction to analysis of simulation output data. *In Proceedings of the 2013 Winter Simulation Conference.*

[50.] Dawid, H. (2006). Agent-based models of innovation and technological change. *Handbook of computational economics*, *2*, 1235-1272.

[51.] de Nooy, W., Mrvar, A., & Batagelj, V. (Eds.). (2005). *Exploratory social network analysis with Pajek* (Vol. 27). Cambridge University Press.

[52.] Drejer, I., & Vinding, A. L. (2006). Organisation, "Anchoring" of Knowledge, and Innovative Activity in Construction. *Construction Management and Economics*, 24(9), 921–931.

[53.] Epstein J. M. & Axtell R. L. (1996). Growing Artificial Societies: Social Science from the Bottom Up (*MIT Press, Cambridge, MA*).

[54.] Epstein, J. M. (2008). Why model?. *Journal of Artificial Societies and Social Simulation*, *11*(4), 12.

[55.] Fishwick, P. A. (1995). Simulation model design and execution: building digital worlds. *Prentice Hall PTR*.

[56.] Fitzgibbons, K., & McNiven, C. (2006). Towards a nanotechnology statistical framework. *In Blue sky indicators conference II* (pp. 25-27).

[57.] Fleming, L., & Frenken, K. (2007). The evolution of inventor networks in the Silicon Valley and Boston regions. *Advances in Complex Systems*, 10(01), 53-71.

[58.] Fujimoto, R. M., Perumalla, K., Park, A., Wu, H., Ammar, M. H., & Riley, G. F. (2003). Large-scale network simulation: how big? how fast?. In Modeling, Analysis and Simulation of Computer Telecommunications Systems, 2003. MASCOTS 2003. *11th IEEE/ACM International Symposium on* (pp. 116-123). IEEE.

[59.] Gaines, B. R. (1979). Sequential fuzzy system identification. *Fuzzy Sets and Systems*, 2(1), 15-24.

[60.] Galaskiewicz, J. and A. Zaheer (1999). Networks of competitive advantage. In S. Andrews and D. Knoke (eds.), *Research in the Sociology of Organizations. JAI Press, Greenwich, CT*, pp. 237-261.

[61.] Garcia, R. (2005). Uses of Agent- Based Modeling in Innovation/New Product Development Research. *Journal of Product Innovation Management*, 22(5), 380-398.

[62.] Gilbert, N., & Troitzsch, K. G. (2005). *Simulation for the social scientist*. McGraw-Hill International.

[63.] Gilbert, N., Ahrweiler, P., & Pyka, A. (2003). The Simulation of Innovation Networks for Policy. *Centre for Research on simulation in social sciences*.

[64.] Gilbert, N., Ahrweiler, P., & Pyka, A. (2007). Learning in innovation networks: Some simulation experiments. *Physica A: Statistical Mechanics and its Applications*, 378(1), 100-109.

[65.] Gittelman, M. (2007). Does geography matter for science-based firms? Epistemic communities and the geography of research and patenting in biotechnology. *Organization Science*, 18(4), 724-741.

[66.] Glahn, H. R., & Ruth, D. P. (2003). The new digital forecast database of the National Weather Service. *Bulletin of the American Meteorological Society*, 84(2), 195-201.

[67.] Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, 51(1), 69-115.

[68.] Glänzel, W. (2002). Coauthorship patterns and trends in the sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies (No. urn: hdl: 123456789/98596). *Katholieke Universiteit Leuven*.

[69.] Glänzel, W., & de Lange, C. (2002). A distributional approach to multinationality measures of international scientific collaboration. *Scientometrics*, 54(1), 75-89.

[70.] Glänzel, W., & Schubert, A. (2001). Double effort= double impact? A critical view at international co-authorship in chemistry. *Scientometrics*, 50(2), 199-214.

[71.] Glänzel, W., & Schubert, A. (2005). Analysing scientific networks through co-authorship. *In Handbook of quantitative science and technology research* (pp. 257-276). Springer Netherlands.

[72.] Glänzel, W., and Winterhager, M. (1992). International Collaboration of Three East European Countries with Germany in the Sciences, 1980-1989. *Scientometrics*, 25(2), 219–227.

[73.] Glänzel, W., Schubert, A., & Czerwon, H. J. (1999). A bibliometric analysis of international scientific cooperation of the European Union (1985–1995). *Scientometrics*, 45(2), 185-202.

[74.] Goldstone, R. L., & Janssen, M. A. (2005). Computational models of collective behavior. *Trends in Cognitive Sciences*, 9(9), 424-430.

[75.] Gould, R. V., & Fernandez, R. M. (1989). Structures of mediation: A formal approach to brokerage in transaction networks. *Sociological methodology*, 89-126.

[76.] Graf, H. (2011). Gatekeepers in regional networks of innovators. *Cambridge Journal of Economics*, *35*(1), 173-198.

[77.] Granello, D. H., & Wheaton, J. E. (2004). Online data collection: Strategies for research. *Journal of Counseling & Development*, 82(4), 387-393.

[78.] Grossman, J. W. (2002). The evolution of the mathematical research collaboration graph. *Congressus Numerantium*, 201-212.

[79.] Guimerà, R., Uzzi, B., Spiro, J., & Amaral, L. a N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science (New York, N.Y.)*, 308(5722), 697–702.

[80.] Gulati, R. (1995). Social Structure and Alliance Formation Patterns: A Longitudinal Analysis. *Administrative science quarterly*, 619-652.

[81.] Gulati, R., Nohria, N. and Zaheer, A. (2000) Strategic Networks, *Strategic Management Journal*, 21(3), 203-215.

[82.] Hagedoorn, J., & Schakenraad, J. (1993). A Comparison of Private and Subsidized R&D Partnerships in the European Information Technology Industry. JCMS: *Journal of Common Market Studies*, 31(3), 373-390.

[83.] Hagedoorn, J., and Duysters, G. (2002). External Sources of Innovative Capabilities: The Preference for Strategic Alliances or Mergers and Acquisitions. *Journal of Management Studies* 39(2):167-88.

[84.] Hao, Z., Yun, X., & Zhang, H. (2006). An efficient routing mechanism in network simulation. *In Proceedings of the 20th Workshop on Principles of Advanced and Distributed Simulation* (pp. 150-157). IEEE Computer Society.

[85.] Hardy, C., Phillips, N., & Lawrence, T. B. (2003). Resources, Knowledge and Influence : The Organizational Effects of Interorganizational Collaboration, 3010(March).

[86.] Heikkinen, M. T., Mainela, T., Still, J., & Tähtinen, J. (2007). Roles for managing in mobile service development nets. *Industrial Marketing Management*, *36*(7), 909-925.

[87.] Helbing, D. (2012). *Social self-organization: Agent-based simulations and experiments to study emergent social behavior*. Springer, ISBN 978-3-642-24003-4.

[88.] Henderson, R., & Cockburn, I. (1996). Scale, scope, and spillovers: the determinants of research productivity in drug discovery. *The Rand journal of economics*, 32-59.

[89.] Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572.
[90.] Huber, F. (2012). Do clusters really matter for innovation practices in Information Technology? Questioning the significance of technological knowledge spillovers. *Journal of Economic Geography*, 12(1), 107-126.
[91.] Hullmann, A., & Meyer, M. (2003). Publications and patents in nanotechnology. *Scientometrics*, *58*(3), 507-527.
[92.] Information Today, 81–93.
[93.] Inzelt, A., Schubert, A., & Schubert, M. (2009). Incremental citation impact due to international co-authorship in Hungarian higher education institutions. *Scientometrics*, 78(1), 37-43.
[94.] Kas, M., Carley, K. M., & Carley, L. R. (2012). Trends in science networks: understanding structures and statistics of scientific networks. *Social network analysis and mining*, 2(2), 169-187.
[95.] Katz, J. S., & Martin, B. R. (1997). What is Research Collaboration?. *Research Policy, 26(1), 1–18.*
[96.] Kollock, P. (1994). The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment, and Trust. *American Journal of Sociology*, 100(2), pp.313-345.
[97.] Kuroiwa, Katsuko (2006). Nanotechnology research in Canada, Retrieved from [http://www.nanowerk.com/spotlight/spotid=984.php#ixzz2TJrqM1Us]
[98.] Lambert, R. (2003). Lambert Review of Business-University Collaboration: Final Report. (2003). *Norwich: HM Treasury.*
[99.] Landry, R., & Amara, N. (1998). The impact of transaction costs on the institutional structuration of collaborative academic research. *Research Policy*, 27(9), 901-913.
[100.] Landry, R., Traore, N. and Godin, B. (1996). An Econometric Analysis of the Effect of Collaboration on Academic Research Productivity, *Kluwer Academic Publishers*, 283–301.
[101.] Lavie, D., & Drori, I. (2012). Collaborating for Knowledge Creation and Application: The Case of Nanotechnology Research Programs. *Organization Science*, 23(3), 704-724.
[102.] Lee, S., & Bozeman, B. (2005). The Impact of Research Collaboration on Scientific Productivity. *Social studies of science*, 35(5), 673-702.
[103.] Lehmann, S., Jackson, A. D., & Lautrup, B. (2006). Measures for measure. Nature, 444, 1003.
[104.] Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology.*

[105.] Lin, S., Cheng, X., & Lv, J. (2008). Micro-Synchronization in Conservative Parallel Network Simulation. *In Principles of Advanced and Distributed Simulation, 2008. PADS'08. 22nd Workshop* on (pp. 195-202). IEEE.

[106.] Lotka, A. (1926). The Frequency Distribution of Scientific Productivity, *Journal of the Washington Academy of Science,* 16: 317-23.

[107.] Macal, C. M., & North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4(3), 151-162.

[108.] Mali, F., Kronegger, L., Doreian, P., & Ferligoj, A. (2012). Dynamic scientific co-authorship networks. *In Models of Science Dynamics* (pp. 195-232).

[109.] Manley, K., Mcfallan, S., & Kajewski, S. (2009). Relationship between Construction Firm Strategies and Innovation Outcomes, (August), 764–771.

[110.] Marsden, P. V. (1982). Brokerage behavior in restricted exchange networks. *Social structure and network analysis*, *7*(4), 341-410.

[111.] Mat, N. A. C., Cheung, Y., & Scheepers, H. (2009). Partner selection: criteria for successful collaborative network. *In 20th Australian Conference on Information Systems* (Vol. 2, No. 4).

[112.] McGinley, S., & Charnie, J. (2003). Peace Building Through Scientific Collaboration. *Arizona University.*

[113.] Melin, G. (2000). Pragmatism and self-organization Research collaboration on the individual level. *Research Policy*, 29, 31–40

[114.] Mervis, J., & Normile, D. (1998). North-South Relations: Lopsided Partnerships Give Way to Real Collaboration. *Science*, 279(5356), 1477.

[115.] Meyer, P. B. (2003). Episodes of Collective Invention. U.S. *Bureau of Labor Statistics*.

[116.] Miyazaki, K., & Islam, N. (2007). Nanotechnology systems of innovation—An analysis of industry and academia research activities. *Technovation*, *27*(11), 661-675.

[117.] Moazami, Afshin. (2012). *A Network Perspective of Nanotechnology Innovation: A Comparison of Quebec, Canada and the United States.* Masters thesis, Concordia University.

[118.] Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion.

[119.] Mowery, D. C., Oxley, J. E., & Silverman, B. S. (1996). Strategic alliances and interfirm knowledge transfer. *Strategic Management Journal*, 17, 77-91.

[120.] Nagpaul, P. S. (1997). Science beyond Institutional Boundaries, New Delhi: National Institute of Science, Technology and Development Studies.

[121.] Nagpaul, P. S. (2002). Visualizing cooperation networks of elite institutions in India. *Scientometrics*, 54(2), 213-228.

[122.] Newman, M. E. (2001a). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2), 404-409.

[123.] Newman, M. E. (2001b). Scientific collaboration networks. I. Network construction and fundamental results. *Physical review E*, 64(1), 016131.
[124.] Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National academy of Sciences of the United States of America,* 101(Suppl 1), 5200-5205.
[125.] Nohria, N., and Garcia-Pont, C. (1991). Global Strategic Alliances and Industry Structure.
[126.] Ocasio, W. (1997). Towards an attention- based view of the firm. *Strategic Management Journal*, 18(S1), 187-206.
[127.] Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6), 441-453.
[128.] Owen-Smith, J., & Powell, W. W. (2004). Knowledge networks as channels and conduits: The effects of spillovers in the Boston biotechnology community. Organization science, 15(1), 5-21.
[129.] Owen-Smith, J., Riccaboni, M., Pammolli, F., & Powell, W. W. (2002). A comparison of US and European university-industry relations in the life sciences. *Management Science*, *48*(1), 24-43.
[130.] Perc, M. (2010). Growth and structure of Slovenia's scientific collaboration network. *Journal of Informetrics*, *4*(4), 475-482.
[131.] Phelps, C., Heidl, R., & Wadhwa, A. (2012). Knowledge, Networks, and Knowledge Networks: A Review and Research Agenda. *Journal of Management* (Vol. 38, pp. 1115–1166).
[132.] Porter, A. L., Youtie, J., Shapira, P., & Schoeneck, D. J. (2008). Refining search terms for nanotechnology. *Journal of Nanoparticle Research*, 10(5), 715-728.
[133.] Powell, W. , & Giannella, E. (2009). Collective Invention and Inventor Networks
[134.] Powell, W. (1996). Inter-Organizational Collaboration in the Biotechnology Industry. *Journal of Institutional and Theoretical Economics,* 120(1):197-215.
[135.] Powell, W., and Brantley, P. (1992). Competitive Cooperation in Biotechnology: Learning Trough Networks?. in R. Ecdes and N. Nohria (eds.), *Networks and Organizations, Boston: Harvard University Press.*
[136.] Powell, W., Koput, K. W., and Smith-Doerr, L. (1996). Interorganizational Collaboration and the Locus of Innovation in Biotechnology, *Administrative Science Quarterly,* 41(1):116-45.
[137.] Price, D. & Beaver, D. (1966). Collaboration in an Invisible College, *American Psychologist*, 21: 1011-18.
[138.] Pyka, A. and P. Saviotti (2000). Innovation networks in the biotechnology-based industries, *SEIN working paper 1*.

[139.] Pyka, A., & Küppers, G. (2002). Innovation networks: theory and practice. *Edward Elgar Pub*.

[140.] Pyka, A., Gilbert, N., & Ahrweiler, P. (2007). Simulating knowledge-generation and distribution processes in innovation collaborations and networks. *Cybernetics and Systems: An International Journal*, 38(7), 667-693.

[141.] Racherla, P., & Hu, C. (2010). A social network perspective of tourism research collaborations. *Annals of Tourism Research*, *37*(4), 1012-1034.

[142.] Rothaermel, F. T., & Alexandre, M. T. (2009). Ambidexterity in technology sourcing: The moderating role of absorptive capacity. *Organization Science*, 20(4), 759-780.

[143.] Sargent, G. (2013). An introduction to verification and validation of simulation models. *In Proceedings of the 2013 Winter Simulation Conference.*

[144.] Sargent, R. G. (2012). Verification and validation of simulation models. *Journal of Simulation*, *7*(1), 12-24.

[145.] Schiffauerova, A., & Beaudry, C. (2008, February). Interaction between geographical and technological spaces of collaboration: The gatekeepers of Canadian biotechnology clusters. In *25th DRUID Celebration Conference on Entrepreneurship and Innovation* (pp. 14-16).

[146.] Schiffauerova, A., & Beaudry, C. (2011). Star scientists and their positions in the Canadian biotechnology network. *Economics of Innovation and New Technology*, *20*(4), 343-366.

[147.] Schilling, M. A., & Phelps, C. C. (2007). Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. *Management Science*, 53(7), 1113-1126.

[148.] Scholz, R., Nokkala, T., Ahrweiler, P., Pyka, A., & Gilbert, N. (2010). The agent-based Nemo model (SKEIN)–simulating european framework programmes. Innovation in complex social Systems. *London: Routledge*, 300-314.

[149.] Schrempf, B. (2013). Project proposal presented at the ManETEI Research Meeting.

[150.] Shan, W., Walker, G., & Kogut, B. (1994). Interfirm cooperation and startup innovation in the biotechnology industry. *Strategic Management Journal*, 15(5), 387-394.

[151.] Singh, J., & Fleming, L. (2009). Lone Inventors as Sources of Breakthroughs: Myth or Reality?. *Management Science*, 56(1), 41–56.

[152.] Soh, P., & Roberts, E. (2003). Networks of Innovators: a Longitudinal Perspective. *Research Policy*, 32(9), 1569–1588.

[153.] Sonnenwald, D. (2007). Scientific collaboration: A synthesis of challenges and strategies. *Annual Review of Information Science and Technology* (41), 643–681.

[154.] Staudt, C. L. (2011). Analysis of Scientific Collaboration Networks: Social Factors, Evolution, and Topical Clustering. Diploma thesis. Karlsruhe Institute of Technology.

[155.] Stuart, T. E. (2000). Interorganizational Alliances and the Performance of Firms: A Study of Growth and Innovation Rates in a High-technology Industry. *Strategic Management Journal*, 21(8), 791-811.

[156.] Thacker, B. H., Doebling, S. W., Hemez, F. M., Anderson, M. C., Pepin, J. E., & Rodriguez, E. A. (2004). Concepts of model verification and validation (No. LA-14167). *Los Alamos National Lab., Los Alamos, NM (US).*

[157.] Tisue, S., & Wilensky, U. (2004). NetLogo: A simple environment for modeling complexity. *International Conference on Complex Systems* (pp. 16-21).

[158.] Tödtling, F. (1999). Innovation Networks, Collective Learning, and Industrial Policy in Regions of Europe, *European Planning Studies*, 7:6, 693-697

[159.] Triulzi, G., Scholz, R., & Pyka, A. (2011). R&D and knowledge dynamics in university-industry relationships in biotech and pharmaceuticals: An agent-based model.

[160.] Tukey, J. W. (1977). Exploratory data analysis. *Reading, Ma*, *231*.

[161.] Van Segbroeck, S., Santos, F. C., Nowé, A., Pacheco, J. M., & Lenaerts, T. (2009). The coevolution of loyalty and cooperation. *In Evolutionary Computation, 2009. CEC'09. IEEE Congress on (pp. 500-505). IEEE.*

[162.] Vaux, J. (2000). Innovation Networks by design, *SEIN Working paper.*

[163.] Velho, L., & Velho, P. (1996). Scientific collaboration of advanced/developing countries in biological sciences: the case of the Maraca Rain Forest Project. *Cadernos de Ciência & Tecnologia*, *13*(1), 9-20.

[164.] Wadhwa, A., & Kotha, S. (2006). Knowledge Creation through External Venturing: Evidence From The Telecommunications Equipment Manufacturing Industry. *Academy of Management Journal*, 49(4), 819-835.

[165.] Wagner, C. S., & Leydesdorff, L. (2005). Mapping the network of global science: comparing international co-authorships from 1990 to 2000. *International Journal of Technology and Globalisation*, *1*(2), 185-208.

[166.] Wagner, C. S., Staheli, L., Silberglitt, R., Wong, A., & Kadtke, J. (2002). Linking effectively: Learning lessons from successful collaboration in science and technology. *RAND CORP SANTA MONICA CA*.

[167.] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *nature*, 393(6684), 440-442.

[168.] Weber, K. M. (2000). The role of networks for innovation diffusion and system change. *CHP in the UK, Germany and the Netherlands* (No. 14). *SEIN Research Report.*

[169.] Weber, M. and S. Paul (1999). Political forces shaping the innovation and diffusion of technologies, *SEIN working paper 4*.

[170.] Weiss, D. M. (1999). Software product-line engineering: a family-based software development process.
[171.] Wetherell, C., Plakans, A., & Wellman, B. (1994). Social networks, kinship, and community in Eastern Europe. *The Journal of Interdisciplinary History*, 24(4), 639-663.
[172.] Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.
[173.] Williamson, O. E. (1983). Credible commitments: Using hostages to support exchange. *The American Economic Review*, 73(4), 519-540.
[174.] Windrum, P. (2000). Innovation Networks in e-commerce, *SEIN Working paper*.
[175.] Wissema, J. G., & Euser, L. (1991). Successful innovation through inter-company networks. *Long Range Planning*, 24(6), 33-39.
[176.] Wooldridge, M. and N. R. Jennings (1995). Intelligent agents: theory and practice, *Knowledge Engineering Review* vol. 10, pp. 115-152.
[177.] Wray, K.B. (2002). The Epistemic Significance of Collaborative Research. *Philosophy of Science*,69, 150-168.
[178.] Yan, E., & Ding, Y. (2011). Discovering author impact: A PageRank perspective. *Information processing & management*, *47*(1), 125-134.
[179.] Yin, L. C., Kretschmer, H., Hanneman, R. A., & Liu, Z. Y. (2006). Connection and stratification in research collaboration: An analysis of the COLLNET network. *Information Processing & Management*, 42(6), 1599-1613.
[180.] Zaheer, A., & George, V. P. (2004). Reach out or reach within? Performance implications of alliances and location in biotechnology. *Managerial and Decision Economics*, 25(6- 7), 437-452.
[181.] Zucker, L. G., & Darby, M. R. (1996). Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. *Proceedings of the National Academy of Sciences*, *93*(23), 12709-12716.
[182.] Zucker, L. G., & Darby, M. R. (2005). Socio-economic impact of nanoscale science: Initial results and nanobank (No. w11181). *National Bureau of Economic Research.*
[183.] Zuckerman, H. (1967). Nobel Laureates in Science: Patterns of Productivity, Collaboration, and Authorship, *American Sociological Review,* 32(3): 391-403.

Appendix I: The Database Dictionary

- **Authors Table**

The information of authors is stored in this table. Each author has a unique id based on the Scopus ID and the followings are the fields of this table:

Name	**DataType**	**Description**
Id	Varchar 50	Author id in Scopus (ex. 35229962500)
FirstName	Varchar 300	First and middle name of the author (ex. Robert E.)
LastName	Varchar 100	Last name of the author
NanoArticles	INT 12	Number of articles by that author which have specialized word in Nanotechnology
AllArticles	INT 12	Number of articles that the author has in Scopus
CitationCount	INT 12	Number of articles that cite this author's articles
hIdex	INT 12	The h Index considersing Scopus articles published after 1995.
CoAuthorCount	INT 12	Number of coauthors of this author based on SCOPUS
Location	Varchar 100	Where the author is currently residing
Firm	Varchar 100	The category of the author`s current affiliation as in 2012
Email	Varchar 500	Author`s email address

- **Articles Table**

The information about the relationships between the authors showing a link between ny two of them if they have coauthored a paper together and the followings are the fields of the table:

Name	**DataType**	**Description**
Authorid	Varchar 50	First author id
Collaborators	Varchar 50	Coauthor id
Paper	Varchar 50	Article id that they have coauthored

- **SNAmeasuerments Table**

The social network analysis measures are collected from Pajek and stored in this table as following:

Name	**DataType**	**Description**
Authorid	Varchar 50	Author id in Scopus (ex. 35229962500)
Betweenness	Decimal 10.7	Betweenness centrality of this node in the network
NorDegree	Decimal 10.7	Normalized degree centrality of this node in the network
Weights	INT 10	Total number of links (partnerships)
Degree	INT 10	Degree centrality of this node in the network (with summed lines)
Weighted_Degree	Decimal 10.7	Connections/Degree
CC	Decimal 10.7	Clustering coefficient of this node in the network

Appendix II: The Survey on Research Partner's Selection

Scientific Collaboration in Canada

* Required

1. **Your primary affiliation is:** *

   Mark only one oval.

   - ◯ Academia
   - ◯ Research Institute
   - ◯ Hospital
   - ◯ Industry
   - ◯ Government
   - ◯ Other: ____________

2. **If you are/have previously been involved in research at university, it is in the following position(s):** *

   Please check all which apply

   Check all that apply.

   - ☐ Graduate Student
   - ☐ Post Doctoral Fellow
   - ☐ Research Assistant
   - ☐ Assistant Professor
   - ☐ Associate Professor
   - ☐ Professor
   - ☐ Not applicable
   - ☐ Other: ____________

3. **Size of the organization that you are currently affiliated to is:** *

   Mark only one oval.

   - ◯ 0-49
   - ◯ 50-99
   - ◯ 100-299
   - ◯ 300-499
   - ◯ 500-999
   - ◯ 1000-4999
   - ◯ >5000

4. **Your primary field of expertise belongs to:** *

   Mark only one oval.

   - ◯ Biology, Life Sciences, Environmental Science
   - ◯ Business, Administration, Finance, Economic
   - ◯ Chemistry and Materials Science
   - ◯ Engineering, Computer Science, Mathematics
   - ◯ Medicine, Pharmacology, Veterinary Science
   - ◯ Physics, Astronomy, Planetary Science
   - ◯ Social Science, Arts, Humanities

5. **You are currently residing in:** *
Mark only one oval.

- Ontario
- Quebec
- British Columbia
- Alberta
- Manitoba
- Saskatchewan
- Nova Scotia
- New Brunswick
- Newfoundland and Labrador
- Prince Edward Island
- Nunavut
- Northwest Territories
- Yukon
- Outside Canada

6. **Your gender is:** *
Mark only one oval.

- Male
- Female

7. **Years of research experience you have:** *
Mark only one oval.

- 0- 5 years
- 5-10years
- > 10 years

8. **Total number of your patents is:** *

9. **Total number of your journal publications is:** *

10. **You are the only author in:** *
How many patents and publications have you done by yourself (no collaboration)

11. **Usually you conduct a research project:** *
Please check all which apply
Check all that apply.

- [ ] Alone
- [ ] With your students only
- [ ] With your supervisor only
- [ ] With an academic partner
- [ ] With an industrial partner
- [ ] With academic & industrial partners for the same project
- [ ] Other:

12. **Indicate the importance for each of the following factors for the selection of your research partner:** *
Mark only one oval per row.

	Unimportant	Slightly Important	Important	Very Important	Critically Important	N/A
The reputation of the organization that research partner is currently affiliated to						
The accessibility to resources, required tools and equipments in research partner's organization						
The availability of funding the research partner is bringing to the project						
Research partner's total number of publications and patents						
Research partner's publications citation rate						
Research partner's career age and years of research experience						
Research partner's reputation in the field						
Research partner has common research area						
Research partner's knowledge in complementary field(s)						
Research partner's personal relationship to you, i.e. friends and family						
Research partner is already within your professional network						
Research partner's prior satisfactory collaboration experience with you						
The strength of the collaboration tie, i.e. the number of your previous common projects and/or the duration of collaboration's relation						
Research partner's geographical location						
Research partner's native language						
Research partner's cultural background						
Research partner's gender						
Research partner's age						

13. **If you want to receive a report summarizing the research findings, please provide your e-mail address:**
Optional

Powered by
Google Drive