Effect of collaboration network structure on knowledge and innovation productivity: The case of biotechnology in Canada

Hamidreza Eslami

Department of Mechanical and Industrial Engineering

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

Master of Science in Industrial Engineering at

Concordia University,

Montreal, Quebec, Canada

December 2011

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: ______________________________

Entitled: ______________________________

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

____

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

Signed by the final examining committee:

______________________________ Chair

______________________________ Examiner

______________________________ Examiner

______________________________ Supervisor

Approved by ______________________________
Chair of Department or Graduate Program Director

____
Dean of Faculty

Date ______________________________

Abstract

Effect of collaboration network structure on knowledge and innovation productivity: The case of biotechnology in Canada

Hamidreza Eslami

Innovation and new knowledge are vital ingredients in establishing and maintaining competitive advantage of companies. Many of the novel ideas that lead to scientific publications or yield innovative output are the result of collaborations among scientists or inventors, who cooperate either on individual level or under organizational agreements. The collaborations that take place among individuals and organizations create a network within which the information exchange occurs. Although various aspects of these networks have been examined, the impact of many network characteristics on knowledge creation and innovation production remains unclear due to the inconsistency of the conclusions from various research studies. One such network structure, called small world, has recently attracted much theoretical attention as it has been suggested that it can enhance the information transmission efficiency among the network actors. However, the existing empirical studies have failed to provide consistent results regarding the effect of small-world network properties on network performance in terms of its scientific and innovative productivity. In this thesis, using the data on 36 years of journal publications in the field of biotechnology in Canada, the network of scientists' collaborations has been constructed based on their co-authorships in scientific articles. Various structural properties of this network have been measured and the level of small-world characteristic has been investigated. We found that the network of biotechnology scientists in Canada exhibits small-world properties. Furthermore, the relationships between these properties and knowledge creation, innovative output and quality of the innovations have been examined. We conclude that the structure of the co-authorship network of Canadian biotechnology scientists has a significant effect on the level of knowledge creation of scientists. However, structural properties of the scientific network have produced impact on neither the quantity nor the quality of innovations produced by the network actors.

Acknowledgements

It is a pleasure to thank those who made this thesis possible. I would like to express my sincere appreciation to my supervisor, Dr. Schiffauerova. This thesis would have not been possible without her guidance and invaluable comments.

I do not think I can ever pay back my parents and my brother for their endless support and encouragement. I would like to express my sincere thanks to my friends for being with me throughout this program. Finally, special thanks go to my dear wife, Dorsa, for her continuous support and patience.

Table of Contents

List of tables viii
List of figures ix
Chapter 1: Introduction and Literature review 1
Introduction 1
Literature review 3
1.1. Network, innovation and innovation networks 3
1.1.1. Network 3
1.1.2. Innovation 3
1.1.3. Innovation networks 4
1.1.3.1. Networks of innovators 6
1.1.3.2. Inter-firm collaborative networks 7
1.2. Structural properties of networks 7
1.2.1. Clustering 10
1.2.2. Path length 12
1.2.3. Interaction of clustering and path length 12
1.3. The notion of small-world and six degrees of separation 14
1.3.1. Regular, random and small-world networks 14
1.4. Effects of network structural properties on its performance 17
1.5. Filling the research gaps 22
Research questions, contributions and objectives 24
Chapter 2: Methodology 27
2.1. Data 28
2.2. Methodology 29
2.2.1. Construction of scientists' co-authorship networks 31
2.2.2. Two-mode and one-mode networks 32
Chapter 3: Analysis of data 36
3.1. Variables 37
3.1.1. Dependent Variables 37
3.1.2. Independent Variables 38
3.1.3. Control variable 46
3.2. Descriptive data analysis 47
Chapter 4: Statistical Analysis 66

4.1. Correlation Analysis 66
4.2. Regression Analysis 69
4.2.1. Regression results for *Total number of articles* model 71
4.2.2. Regression results for the *Total number of patents* model 73
4.2.3. Regression results for the *Total number of patent claims* model................ 74
Chapter 5: Conclusions, Limitations and Future Study 76
5.1. Conclusions 76
5.2. Limitations of the study 79
5.3. Future study 80
References 83
Appendix A: Examples of two-mode and one-mode networks of Canadian biotechnology scientists 91
Appendix B: Results of the likelihood-ratio test for the over dispersion coefficient (Alpha) 93

List of tables

Table 1: Data summary 47

Table 2: Descriptive statistics of the variables 49

Table 3: Small-World characteristics for the Canadian biotechnology networks 64

Table 4: A comparison of previously studied small-world networks 65

Table 5: Correlation analysis of Canadian biotechnology network 67

Table 6: Poisson regression results for *total number of articles* model............................ 71

Table 7: Poisson regression results for total number of patents model 74

Table 8: Negative binomial regression results for the total number of patent claims model 75

Table 9: Likelihood-ratio test for over dispersion coefficient of articles model 93

Table 10: Likelihood-ratio test for over dispersion coefficient of patents model............. 93

Table 11: Likelihood-ratio test for over dispersion coefficient of patent quality model .. 94

List of figures

Figure 1: From a regular network to a random one (Watts and Strogatz 1998) 15

Figure 2: Changes of clustering coefficient and path length by variation of the number of random links in the network, Schilling and Phelps (2005) 16

Figure 3: Two-mode affiliation network of article co-authorships 35

Figure 4: One-mode network of scientists 35

Figure 5: Historical trend of network size from 1970 to 2005 50

Figure 6: Historical trend of network connectivity from 1970 to 2005 51

Figure 7: Historical trend of closeness centralization from 1970 to 2005 52

Figure 8: Historical trend of the number of scientists in the largest connected component of Canadian biotechnology network between 1970 and 2005 54

Figure 9: Historical trend of the proportion of the scientists in the largest component in Canadian biotechnology network between 1970 and 2005 54

Figure 10: Historical trend of degree centralization of Canadian biotechnology network from 1970 to 2005 55

Figure 11: Historical trend of betweenness centralization of Canadian biotechnology network between 1970 and 2005 56

Figure 12: Historical trend of actual values of path length of the Canadian biotechnology network from 1970 to 2005 57

Figure 13: Historical trend of clustering coefficient of the Canadian biotechnology network from 1970 to 2005 59

Figure 14: Path length ratio of Canadian biotechnology number (based on the Watts method) 62

Figure 15: Clustering coefficient ratio of Canadian biotechnology number (based on the Watts method) 62

Figure 16: Two-mode affiliation network of Canadian biotechnology scientists from 1969 to 1973 91

Figure 17: One-mode network of Canadian biotechnology scientists from 1969 to 1973 92

Chapter 1: Introduction and Literature review

Introduction

Although the concept of networks has been known for many years, it is a short time that human being has started to study its features and characteristics. The benefits of networking have been proven in various fields, and physicians, mathematicians, and others have been utilizing them for some time. However, its application in business world, especially in enhancing the innovation and creativity, has only recently attracted some theoretical attention.

Networks in which various firms and individuals have relationships with each other are of great importance. Especially the relations involving the exchange of knowledge play a key role in the survival and the progress of organizations. In fact, the knowledge transfer is one of the major factors that bring innovative companies close to each other and shape the geographical clusters. The effective flow of knowledge in the networks will result in the regional improvement of knowledge level and lead to economic growth.

The more connections to other firms a company has and the more valuable knowledge it gains through these relationships, the more successful it will be in creating innovation and developing new products and procedures. In any network, as the population of nodes grows, the number of connections increases as well. In the large social networks, as soon as some knowledge or information is available, it can spread to all parts of the networks due to the presence of direct and indirect links. Hence any distance between the innovative agents can affect the extent of knowledge diffusion strongly. This fact eventually influences the performance of the agents in the network and also their innovative productivity.

Therefore, we can realize the importance of studying the structure of social networks and the process of knowledge and innovation diffusion. In this thesis, the network structure and its properties will be studied and some light will be shed on their effect on innovation productivity.

Since most of the research work carried out in this field is theoretic, there is a vivid lack for empirical investigations based on statistical data. The data used in this study are taken from the field of biotechnology in Canada. This segment of industry is of great importance for Canada, because it is a relatively new sector with a great potential for growth. Furthermore, biotechnology provides a significant contribution to science advancement and innovation, thousands of jobs, as well as large exports.

The content of this thesis will provide a deeper understanding of the effect of the innovation network's structure on the knowledge creation and innovation productivity level. As such, it can serve as a basis for the design of governmental policies or organizational strategies related to knowledge creation and its transmission through the networks. It is proposed in this thesis that efficiently structured knowledge networks can finally result in the increase of knowledge productivity and innovativeness in Canadian biotechnology sector.

Literature review

1.1. Network, innovation and innovation networks

1.1.1. Network

Generally, a network consists of nodes or points which are connected to each other by links. In industrial societies firms and organizations can be thought of as nodes of the network and any type of relationship between them is considered as links that connect the network nodes to each other. This fact holds also for individuals in companies, universities and other institutions, which means, for example, that we can build a network of innovative individuals like scientists and inventors who are related to one another by their cooperation and co-authorship ties.

The connection of the nodes varies in different networks and therefore the networks exhibit diverse characteristics. However, according to Wasserman and Faust (1995), in the networks of innovators and firms, there are regularities in their relationship which shape some patterns. These patterns are known as the structure of the network and will be discussed later.

1.1.2. Innovation

It is needless to say that in any industry, innovation improves the companies' value; it helps them maintain their competitive advantage and enables them to enter into new markets. Besides, innovation enhances the knowledge level in various fields. Not only firms and individuals share their knowledge, information and achievements with some other agents through the networks, but also many organizations invest in their own research activities and inventive personnel in order to successfully compete with others. This becomes a strong motive for them to improve their innovative performance. Now, we should define what is called innovation:

According to Dosi 1998, the procedure of search to find the solutions for problems is called innovation. It is believed that the produced knowledge usually is the result of the new mixture of existing information and solutions; or it is generated from the new composition of knowledge components (Schilling and Phelps 2007).

1.1.3. Innovation networks

In today's competitive world, organizations prefer to collaborate with each other to take advantage of the knowledge trade among themselves. Although some people still believe that it is better to work alone and not to share their ideas, recent research has shown that most of novel inventions and scientific achievements have been derived from collaborations and partnerships. For example, Collins (1999) analyzed the business line of many remarkable geniuses among artists, scientists and philosophers, and found that majority of famous people (for example Freud and Beethoven) worked in connection with others who were in fact often their rivals and competitors.

There are various kinds of connections among knowledge sources, but in general, we recognize direct and indirect collaborative ties. In direct collaborations, two knowledge sources (organizations or individuals) have direct connections between each other, which are based directly on their physical collaboration. However, in indirect collaborations, there are intermediaries between the two (or more) agents that exchange knowledge and information, and knowledge is thus transmitted through the complex net of links and relationships.

It has been shown that this network of agents greatly facilitates the knowledge transfer and diffusion (Schilling and Phelps 2007). There could exist a very rich and fruitful exchange of knowledge and innovative ideas among the members of such social networks. These networks can involve various research fields in which they can significantly contribute to the development of new technologies, new medicines or many other kinds of innovation (Abrahamson and Rosenkopf 1997). Previous studies show that firms' relationships affect their achievements in innovation and even indirect connections increase the agents' innovative performance (Schilling and Phelps 2007).

The abovementioned relationships could develop among different sources of knowledge. According to Midgley *et al.* (1992), there are eight categories between which the knowledge exchange takes place in organizational level. These are as follows:

1. Suppliers
2. Adopting organizations
3. Adopting organizations and suppliers
4. Government regulations and adopting organizations
5. Government regulations and suppliers
6. Suppliers and adopting organizations
7. Other third parties and adopting organizations and vice versa
8. Suppliers and Other third parties and vice versa

Allen (1983) believes that the free flow of information among these agents leads to the formation of collective inventions. In fact this phenomenon is the result of the sharing of knowledge among groups and will not happen due to the attempts of neither individuals nor particular organizations. In the case that Allen (1983) has studied, the knowledge circulation within the firms in blast furnace industry led to an outstanding advance in the performance of blast furnaces in Britain. There are many other instances that confirm the positive effects of the disclosure of knowledge within the connected social groups of individuals and/or firms that form the networks. According to Schiffauerova and Beaudry (2008a), collective invention leads to the rapid growth of knowledge in the network and increases the innovation production rate.

Based on subjects under the study we classify the innovation networks into two main categories: the networks of individual innovators, i.e. inventors or scientists, and the inter-organizational networks composed of the firms and organizations.

1.1.3.1. Networks of innovators

A large number of inventors and scientists cooperate with each other, facilitate thus knowledge transfer within their communities, and collectively contribute to the generation of new scientific achievements and the production of innovations. Their connections form an inter-personal network of researchers and inventors, which we call the network of innovators. Although these individuals are frequently cooperating, they usually do not have any official contracts evidencing their collaborations. Their relationships are often traced through the results of their scientific and innovative efforts through scientific articles and patents. Therefore, their association is built on two factors:

co-inventorship of a patent and co-authorship of an article. The former can be tracked by the patent documents and the latter is evidenced by the scientific journals. According to Newman (2001a) the co-authorship networks built in this way belong among the largest social networks ever studied.

1.1.3.2. Inter-firm collaborative networks

The second type of innovation networks is formed among the firms rather than individuals. In fact, the collaborative links of firms with other organizations develop knowledge networks whose network structure and properties are different from networks of innovators. The evidence of inter-firm partnerships is based on various data, for example on officially registered alliances, collaborative research agreements and also joint ownership of patents (Schiffauerova, A., Beaudry, C. 2009).

1.2. Structural properties of networks

No one can ignore the economic and social importance of networks connecting different organizations and individuals. The development and the spread of innovations, their success rate and also the innovative potential of the firms are highly affected by the structure of the network over which communication takes place (Midgley *et al.* 1992). It has been proposed that the successful spread of knowledge or invention is dependent on the structure of the network in which it flows. Even for highly valuable innovations, only their introduction to the social network's section is not sufficient for its successful diffusion, and many other factors affect its success (Abrahamson and Rosenkopf 1997).

Although the network structure in which the vertices (agents) are connected and exchange knowledge, has not been much explored in research literature, the network architecture is considered to be a crucial factor that influences the type of transferred knowledge as well as its amount and the transfer effectiveness (Cowan and Jonard (2004). The information transfer among network members affects the knowledge productivity and innovative performance of the network; hence it could be concluded that network structure is a significant factor in the improvement of network's knowledge level.

Much evidence supports the fact that some properties of networks influence the spread of knowledge. For example, in the survey done by Abrahamson and Rosenkopf (1997), a scientist named Dr. W. Edwards could not spread his new approach (TQM) in the U.S. because he was not famous there. Therefore, he went to Ichiro Ichikawa and through this well-known scholar in Japan he diffused his approach to many Japanese and then to U.S. segments. In the same study, it was also described how James Lancaster and James Lind's findings about lime juice cure property for scurvy was ignored because of their little stature in British navy social networks.

Schiling and Phelps (2007) have also mentioned the important role of the structure of networks connecting firms and its significant effect on the flow of knowledge among these companies. Many other researchers like Cowan and Jonard (2001), Abrahamon and Rosenkopf (1997), Choi *et al.* (2010) and Granovetter (1973) have emphasized the outstanding effect of network topologies on the performance of the system and the diffusion of knowledge and innovation. Considering the importance of these effects it is recommended to pay great attention to the architecture of innovation networks. The first step in this regard taken in this thesis is to get acquainted with properties of networks over which the communication takes place. The network properties that enhance the knowledge transfer as well as those which reduce the rate of information diffusion should be studied and well understood.

Generally, each social network is made up of some internal segments which are separated based on geographical properties, cultural properties or industry types. These segments form boundaries which may prevent innovation to spread to all of the potential adaptors by limiting the diffusion process. Therefore, internal segments of networks could greatly affect the extent of knowledge diffusion. (Abrahamson and Rosenkopf 1997)

Some scholars have surveyed the various properties of networks which influence the spread of knowledge. For instance, Granovetter (1973) mentions that two main constituents that form each social network are “cliquish sub-networks” and “bridges”. Cliques, such as family or friend networks, consist of people or firms who widely cooperate. He says that the information of such sub-networks will be diffused within a few cliques. But bridge connections tie many agents from various cliques. A simple example could be international conferences or internet chat rooms.

In the current research literature, two major structural features of networks that are recognized to have an important effect on the network performance are called clustering (also known as cliquishness) and reach (also known as path length) (Schilling and Phelps 2007, Watts and Strogatz 1998, Granovetter 1973). These characteristics are described below.

1.2.1. Clustering

Clustering coefficient is defined as "the proportion of a firm's partners that are themselves directly linked to each other". For the whole network, the clustering coefficient can be gained by averaging the clustering coefficients of all agents (Schilling and Phelps 2007). Watts and Strogatz (1998) introduce clustering coefficient C(*p*) as the local cliquishness[1] of the network. C(*p*) is defined as the likelihood that two nodes are connected in case that they are both connected to a mutual node. Basically, each three people in a network who are connected to each other make a triad. This triad is a sample of complete clustering in which C(p)=1. For any network, clustering coefficient is calculated by computing the ratio of total number of triads in the network over the number of all possible triads (Uzzi 2008).

According to Schilling and Phelps (2007), clusters could be shaped due to many reasons, but commonly, proximity of the organizations and also their similarity lead to high clustering. For example firms which are geographically close or firms that have similar technologies are more prone to communicate with each other.

Schilling and Phelps (2007) claim that clustering leads to higher knowledge diffusion capacity and network performance. The reason is that with higher clustering there will be more alternatives of solutions for existing problems across the entire network, and this increases the general insight of the members. Besides, when the network is more clustered, there will be more trust among its members. Consequently, enhanced trust

[1] Cliquishness is another term used by many scholars to refer to clustering. (Burt 2001, Uzzi and Spiro 2005 and Schilling and Phelps 2007, Cowan and Jonard 2003, Fleming et al. 2007)

within the dense lattices improves the cooperation among the firms. Along with more collaboration and reciprocity, staff of enterprises will be motivated to exchange knowledge with other agencies' personnel. Furthermore, the denser a cluster is the quicker will be the transmission of shared knowledge. Moreover, even the number of clusters affects knowledge distribution and productivity of the network; when there are many clusters, the dispersed information is more probable to contain various fields of knowledge. This will enhance the overall knowledge level in the network.

Nevertheless, there are also some drawbacks associated with higher levels of clustering. As an example, when the cluster becomes denser, the amount of information which flows among the cluster members will increase. Consequently, even though the level of the diversity of transferred knowledge will increase, the extent of information which is identical and redundant may increase as well (Burt 1992, Granovetter 1973). The link redundancy is characterized by many unneeded links to the same sources of knowledge. In agreement with this phenomenon, Cowan and Jonard (2004) have concluded that when two agents are in the same clustered clique, obviously there will be many paths by which the information and innovative material can be transmitted. Although this will increase the innovation productivity of the members, there will be many redundant connections which result in the exchange of identical information

Moreover, Uzzi and Spiro (2005) also suggest that although clustering is considered as an important factor in many innovation systems, it may have negative effect if it exceeds an acceptable level. They have stated that when the cliquishness passes a certain threshold, diffusion properties become weak in the network. Consequently, with a drop in the extent of knowledge spread among network nodes, the innovation and knowledge productivity of the network will decrease.

1.2.2. Path length

According to Watts and Strogatz (1998), the average number of edges that should be traversed in the shortest path between any pair of vertices is called characteristic path length L(*p*) and determines the separation between two vertices in the network. Actually it is "the average number of links that separates each pair of firms in a network" (Schilling and Phelps 2007).

Short path length with many knowledge sources makes the access to more information possible. The path length between two agents in a network affects the possibility of knowledge exchange between these agents and also the speed by which they are able to exchange the knowledge. Obviously, the shorter is the average path length between a certain individual or firm and other network actors, the more knowledge can reach that agent (individual or firm) quickly (Watts 1999, Uzzi and Spiro 2005, Schilling and Phelps 2007).

1.2.3. Interaction of clustering and path length

Although it was previously assumed that there should be a balance between high clustering and many short path lengths in a network (Schilling and Phelps 2007), recently it has been proposed that even a few connections among agents can provide them with distinct knowledge fields from different clusters (Fleming *et al.* 2007, Uzzi and Spiro 2005, Schilling and Phelps 2007).

Study results of Schilling and Phelps (2007) show that the interaction of high clustering and short path length has positive influence on the productivity of knowledge. In their model, a small increase in clustering enhances the positive effects of short path length on knowledge creation.

Simulation results of Choi *et al.* (2010) demonstrate that the low degrees of cliquishness and high randomness[2] (which is the result of high number of bridges[3]) reduce the complete spread of information and innovation throughout the network. In their model, when the number of random connections increases, the likelihood of failure in complete diffusion will raise in its early stages of the introduction of the new knowledge to the network. Whereas when randomness is low, the adoption of new ideas is high in this period (The period in which the ideas are newly introduced).

Consequently, in initial stages of introduction of innovation to the network (especially when the innovation is related to new product), a cliquish network shows more diffusion than a random one; however, when this stage has passed, random connections make the diffusion faster. This fact does not hold for information diffusion, meaning that high randomness of the network leads to acceleration of information diffusion speed. For product adoption, there should be a balance between high randomness and high cliquishness. This could happen using a cliquish network with some random links which refers to a special network structure named small-world.

[2] When the randomness increases in a network, its disorderness will be raised. In this case the likelihood that two close nodes are connected is much lower than in the regular networks.

[3] Bridge refers to the links that connect different clusters to each other in a network. According to Schilling and Phelps (2007) bridges enable the network members to reach many sources of knowledge. Uzzi and Spiro (2005) have suggested that the existence of bridges increases the chance of having access to various ideas, which leads to the higher likelihood of knowledge recombination and therefore it enhances the creativity and the productivity of knowledge.

1.3. The notion of small-world and six degrees of separation

When we find a mutual acquaintance with someone whom we do not know at all, it brings a thought to our mind that "the world is really small". This means that any two people around the world who are randomly selected are connected to each other with some intermediate links.

Milgram (1967) was the first one who did a quantitative survey regarding the small-world notion. He randomly selected 296 individuals in Nebraska and gave them letters to be delivered to a specific person in Boston whom they did not know. They were asked to use their acquaintances which would pass the letter further and further, and have it finally delivered to the addressee. The results of this survey show that on average only six intermediates were needed to reach the person who was completely unknown to those individuals sending the letter. This study concluded that each pair of people in the world is separated on average by six intermediate acquaintances.

Later, this phenomenon was named "six degrees of separation" (Gaure 1990). After that, Watts and Strogatz (1998) introduced a model of small-world in which there are some clusters that contain local ties among agents and also there is a few global links that enable connections between any pair of nodes in the network.

1.3.1. Regular, random and small-world networks

Generally, all the network connections are believed to be either regular or random. However, between these two extremes of networks (completely regular or completely random) there could be many social, biological or technological lattices. Small-world network structure, which falls between these two extremes, has resulted from enhanced amount of disorder in regular graphs by rewiring them. In small-world networks high clustering could coexist with short path lengths (Watts and Strogatz 1998).

To construct the network structure falling between the two mentioned extremes, Watts and Strogatz (1998) rewired each link with probability p. In this case, p=0 makes the graph with complete regularity and p=1 leads to a random (disordered) graph. (See Figure 1)

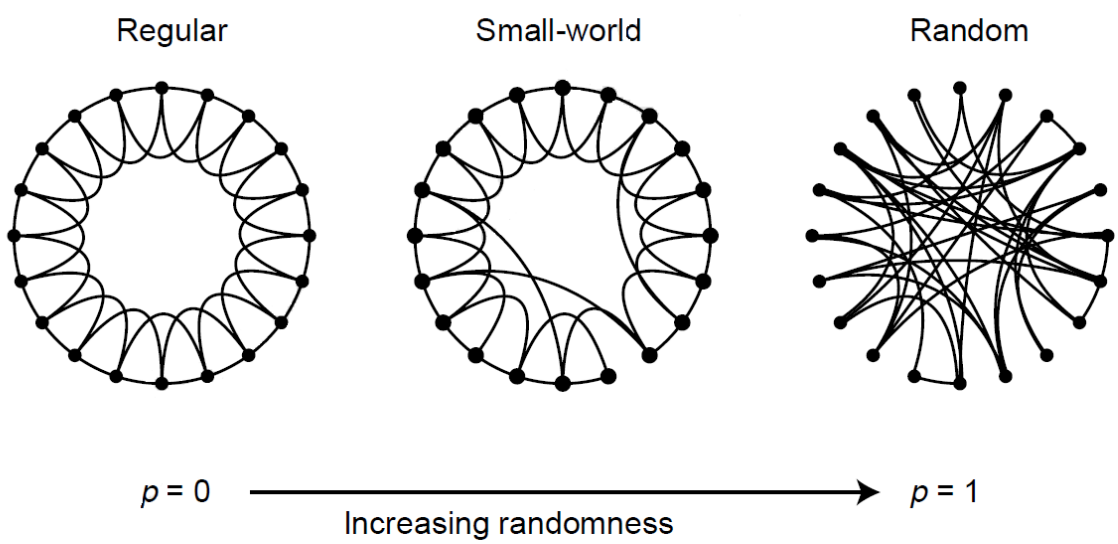

Figure 1: From a regular network to a random one (Watts and Strogatz 1998)

In regular networks, the agents are directly connected to their closest neighbors, but the paths between the nodes located far from each other involve many indirect links. Therefore, in this kind of networks path length is long and also clustering is high. In random networks, members of the network are connected randomly to each other. So the probability that two close neighbors are connected is much less than in the regular lattices. Consequently, these networks are characterized by short path length and low clustering. Between these two extreme kinds of networks, there are small-world graphs in which high clustering and short path length exist concurrently. Thus, small-world networks have some features of random graphs (short path length) and some of regular ones (high clustering) (Fleming *et al.* 2007). The figure below demonstrates that by increasing the number of random links both the clustering coefficient and the average path length in the network decreases, but the path length decreases much faster. This creates an interval in which high clustering and short path length coexists and in which the properties small-world networks are found.

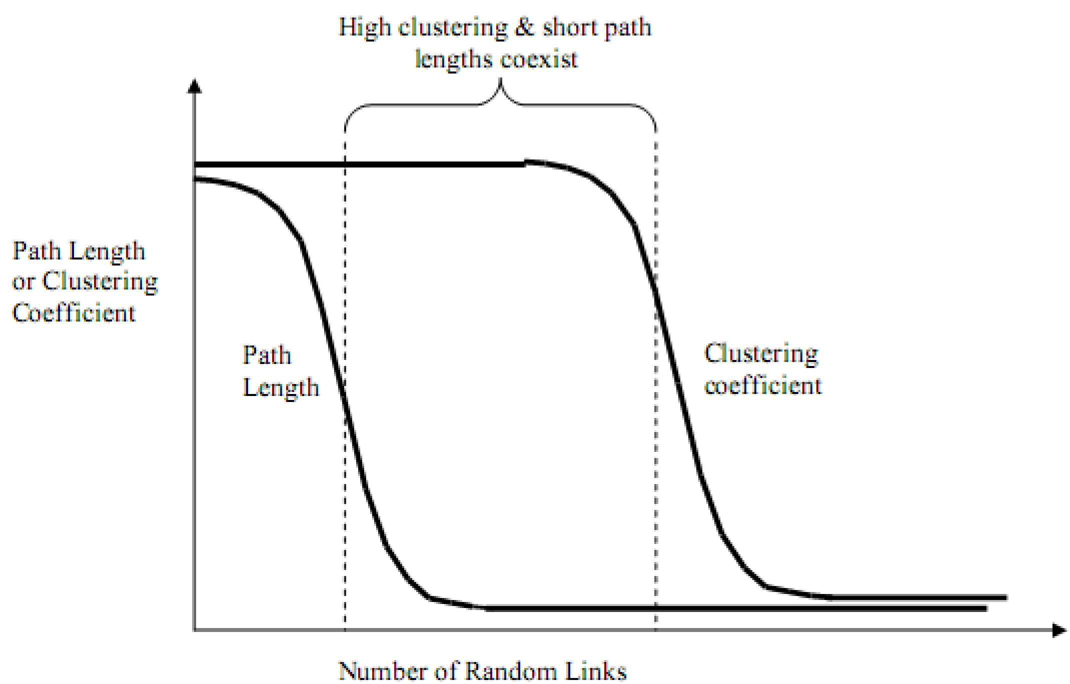

Figure 2: Changes of clustering coefficient and path length by variation of the number of random links in the network, Schilling and Phelps (2005)

1.4. Effects of network structural properties on its performance

Many research studies have assessed different aspects of network structural properties (and especially small-world characteristics). For example, Travers and Milgram (1969) tried to formulate the small-world by calculating the probability of any two randomly chosen people knowing each other in a large population. The study was performed in America and the observed mean of intermediaries, which is around five, is proved to be stable.

In the work of Choi *et al.* (2010) which studies the effects of network structure on the innovation diffusion, it is proposed that the probability of distribution of a new product is higher in random lattices than in the dense ones. They believe that for an innovation in early phases, network randomness makes it difficult to take advantage of network effects and this will prevent innovation to be diffused completely. However, when the diffusion progression reaches a certain step, the randomness of network leads to a faster spread of knowledge.

Latora and Marchiori (2001) have evaluated the efficiency of some specific networks (neutral networks, communication networks, and transport networks), by comparing them to small-world networks as globally and locally efficient networks. They define the efficiency of the networks by the efficiency of information exchange in it.

The efficiency measure introduced in the survey of Latora and Marchiori (2001) enables them to generate a clear physical meaning to small-world, and quantitatively analyze the information flow in various networks. They imply that this measure is applicable in both un-weighted and weighted networks, and it can be used both in theoretical and empirical cases. Comparing to the real data, they conclude that various existing networks (neural, communication, and transport) are similar to small-world networks, and are therefore globally and locally efficient.

Some other aspects of the efficiency of small-world networks have been discussed by Cowan and Jonard (2004). They have studied the effect of network architecture on the performance of the diffusion. They claim that the level of knowledge is at its maximum when the network structure has the small-world properties. They define the small-world lattice as the one in which the number of links connecting a vertex to the other vertices which are outside of its neighborhood is between 1 to 10 percent of all the existing direct links in the network.

Cowan and Jonard (2004) have developed a model in which knowledge exchange among agents occurs only when is it mutually profitable, i.e. when it increases knowledge levels of both actors. They have varied the randomness level in the network and measured the mean knowledge level over the whole network as the performance measure. Their studies show that the networks with "small-world" properties have higher mean knowledge level.

Some scholars suggest that enough theoretical studies are available about the small-world network and it is time to practice these theories through empirical studies. In this regard Fleming *et al.* (2007) have tried to empirically approve or reject some of the hypotheses about small-world networks, such as the improvement of innovative creativity in the presence of small-world effect, the improvement of collaboration and trust between close firms in the networks, and the effect of distant connections between clusters in bringing new knowledge to the clusters. They chose US regions to execute their empirical study while using the networks built on the patent co-authorship data.

Fleming *et al.* (2007) have tried to shed some light on the effects of small-world on the inventor networks and their innovative and managerial approaches within a small-world network to remain competitive. Their results support the positive influence of short path length on innovative productivity. However, their study failed to show that the small-world structure have significant positive influence on the innovative performance of the network.

In another study by Sullivan and Tang (2010) the inter-firm network of United States venture capital industry has been constructed to evaluate its effects on the firms' performance. They concluded that the productivity of firms is improved by the small-world properties. However, the various abilities of firms in absorbing new information have distinct effects on the productivity improvements associated with small-world characteristics. Moreover, Kogut and Walker (2001) conducted a study on the Canadian network of investment bank syndicate from 1952 to 1990 to see how small-world network emerges and evolves over time. They tested and confirmed the hypothesis that the networks formed among firms usually resemble small-world characteristics.

He and Fallah (2006) performed a case study on the inter-organizational level of collaborations in Texas and New Jersey. They built the networks based on the patent co-authorship relationships to investigate the innovative productivity of the networks. They used number of patents as an index for innovativeness of the organizations as a function of structural properties of the network. Their main focus of their study was on the connectivity of network and centrality of the nodes. Their comparisons between the networks of the two regions suggested that the central companies (like Bell systems) in New Jersey play an important role in the innovation network of region; whereas in the network of Texas, the role of central companies is much less effective.

McFadyen et al. (2009) considered the individual level of collaborations in innovation networks of the scientists based on article co-authorships of university researchers. They evaluated the knowledge creation of scientists as a function of the properties of the networks among them. Their study covers eleven years of collaboration data, and explored the strength of ties among scientists as well as the structural properties of their network. They concluded that both of the aspects they measured for the researchers' network (structural properties and tie strength) affect the knowledge creation of the network. They claimed that the scholars, who make strong relationships with other individuals who themselves do not have many collaborators, have high knowledge creativity.

Newman (2001a, 2001b, 2001c, 2004) has carried out several research studies examining the network among the individual scientists with an aim to analyze different structural properties of the network, including small-world effect. He has studied the article co-authorship network in physics and biology to find the effect of number of common collaborators between scientists on the probability of their own collaboration. He also explored the probability of having new collaborators for scientists based on their number of past collaborations (Newman, 2001c). In another study (Neman 2004), by using the the same approach (researcher's articles co-authorship) he built the networks of three scientific fields, biology, mathematics, and physics, and compared some structural properties of these networks, such as the number of publications, distance between mutual scientists, and the clustering of scientists' network to each other.

1.5. Filling the research gaps

Based on the interest and the attention that social networks have received in recent years, it is evident that the subject represents an important, interesting and fruitful approach for the study of social systems. Moreover, the wide variety of topics in this area and their application in different fields create a vast field for conducting future research.

After the introduction of the first notion related to the small-world networks by Milgram (1967), this subject has been investigated by many researchers in various fields. Many empirical studies have been performed in different contexts to analyze the small-world effect in social networks including German corporate ownership, American corporate boards, strategic alliances, Canadian investment bank syndicates, email networks, Italian scientific and academic collaboration networks, and invisible scientific colleges. (Kogut and Walker 2001, Davis *et al.* 2003, Verspagen and Duysters 2003, Baum *et al.* 2003, Dodds *et al.* 2003, Balconi *et al.* 2004, Goyal *et al.* 2004). As it is assumed that the level of the influence of network properties is different in distinct industries (Felmand and Andretsch 1999), more studies elucidating this impact for various industrial sectors are needed. The effect of collaboration in Canadian biotechnology has not been considered so far and the present thesis is going to fill this research gap.

The effects of the small-world network structure have already been also investigated in the field of innovation. It has been proposed that the small-world network properties can have an immense effect on enhancing the knowledge and innovation production (Watts 1999, Hargadon 2003, Cowan and Jonard 2003, Baum et al. 2003, Schilling and Phelps 2007, Uzzi and Spiro 2006). However, despite the significance of this area under discussion, as mentioned before, the amount of practical and empirical research performed within this theme is still scarce (Fleming *et al.* 2007). This thesis will contribute to the research literature by providing an empirical study on the impacts of the network structure effects on network performance in the field of biotechnology.

Moreover, as for the impact of properties of social network structure on the extent of innovativeness and knowledge productivity, the conclusions of various research studies are not very consistent. As an example, the results of various articles on the impact of clustering on the innovative performance are fairly distinct and often even support quite opposite effects. It is thus one of the objectives of this thesis to shed some light on the impacts of various network properties on the scientific and innovative productivity of the networks.

Furthermore, all the discussed research work focuses on investigating the network structure effects either within the realm of the scientific academic networks or within the industrial innovation network. This thesis is, to our knowledge, the first attempt to bring those two worlds together and to analyze the effect of the network of scientific collaboration on their subsequent innovative productivity.

Research questions, contributions and objectives

Generally, this project addresses five main research questions. First, we will examine to what extent the structure of the collaboration network of Canadian biotechnology scientists resembles the small-world network structure. Since this type of network structure has become of interest of many researchers recently, and its positive effect and optimality is still under question, it is important to find out if the network under study shows small-world characteristics.

Second, the effect of the structure of collaboration network of Canadian scientists in the field of biotechnology on their research productivity is examined. We are interested to test various structural prosperities of the network and see their impact on the creation of knowledge by scientists. Some other scholars have asked the same question, but since their results are not consistent and the outcome could be different for distinct technology sectors, we have tackled this question.

Third, we will investigate whether the structure of the Canadian scientists' co-authorship network affects the level of innovation productivity of Canada in the field of biotechnology. In this regard, this study will provide statistical evidence to evaluate the role of network structure by quantifying the properties of scientists' network.

Four, in this study, we will determine the influence, if there is any, of scientists' collaboration network on the quality of innovative output. The quality of publications has not been examined as widely as their quantity, and since we may get dissimilar results from different networks, this research question merits more investigation for the biotechnology industry in Canada.

The last important question we posed is: does the small-world structure (if exists) facilitate the knowledge creation and the innovative performance of the inventors? Previously, it has been widely accepted by many scholars (For example: Cowan and Jonard 2004, Schiling and Phelps 2007, Watts 1999, Hargadon 2003) that the small-world structure enhances the innovative productivity of the inventors' network to a great extent. The effect of this kind of network structure is investigated on the quality of the innovation output of the inventors as well as on the research productivity of scientists in the Canadian sector of biotechnology scientists.

This research represents a longitudinal study on the network of Canadian scientists. Compared to previous studies it covers a very long period of time (1966 – 2006). Furthermore, most of the prior researchers analyzed only the effect of patent co-inventorship networks on the innovation productivity of inventors; but this study takes one step further by taking into account the important role of scientists' reciprocal knowledge transfers during the creation of their scientific knowledge (represented here by the article co-authorships), in promoting the innovativeness of biotechnology scientists.

In summary, this study has several different objectives. First of all, it explores the structure of the network of relationships between Canadian scientists in biotechnology by measuring many of its properties. It also investigates the extent to which the Canadian scientists' network corresponds to the small-world structure.

Moreover, this thesis attempts to discover the impact of network structure on the innovation productivity, the innovation quality and also on the research productivity of individual scientists and inventors. The results of this study are of great importance because all the analyses are based on a real data taken from biotechnology publications and patents.

Chapter 2: Methodology

In this chapter, the methodology by which the research questions are examined is explained in detail. As mentioned before, this study will evaluate the network of Canadian biotechnology scientists. It has been shown (Powell 1990) that the network ties existing among the researchers, inventors, universities and companies are among the most important factors that move the biotechnology industry forward and that they have a significant effect on the knowledge productivity. This also explains the recent interest of many scholars attempting to assess different aspects of biotechnology innovation in adopting the network analysis point of view.

Furthermore, one of the major attributes of biotechnology sector is the dominance of scientists' social network (Oliver 1996). The scientists working on the common research projects and publishing scientific articles together exchange a great amount of biotechnology specific knowledge among themselves (Demirkan 2007). Therefore, the collaboration networks of these scientists are considered to be significant drivers of research progress. Many of the information exchanges and knowledge transfers take place at the individual level (among scientists or inventors) either within the companies, universities, governmental laboratories or among the individuals from different organizations (Oliver & Liebeskind 1998, Liebeskind 1996, Zucker et. al 1995).

It is thus rather surprising that the majority of prior research studies considered only the inter-organizational relationships and analyzed the networks of the alliances or partnerships among the companies, universities and other institutions, while the collaboration at the individual level has been neglected. This study focuses on this research gap and performs an empirical analysis of the collaboration network of individuals who perform a scientific research in biotechnology and who are affiliated to the research institutes, governmental institutes, firms and universities located in Canada.

2.1. Data

The abovementioned networks were built based on the existing databases. These databases include the data on the scientific articles extracted from the SCOPUS database and the patents extracted from the United States Patents and Trademarks Office database (USPTO) database. Various sources of patent data were considered, for example Canadian Intellectual Property Office database (CIPO) and European Patent Office (EPO). In this study the data from USPTO has been used in which, unlike the other databases, the geographical location of the inventors is provided. This may cause a certain bias, but according to Schiffauerova and Beaudry (2008a), most of the Canadian biotechnology inventors prefer to protect their intellectual properties in US. This is due to different reasons. For instance, the biotechnology market of United States is larger; and the access to it is much easier than the Canadian market. Besides, due to the long development cycles of biotechnology products, the accessibility to a large market is of great importance for inventors in order to have satisfactory returns of investment. The USPTO database is thus considered as a suitable source of data for assessing the innovation of Canadian biotechnology sector.

To build the network of scientists (article authors) and analyze the impact of its structure on knowledge and innovation, we also need the data on all publications related to the field of biotechnology in Canada. Here, the SCOPUS database has been selected because it covers a significant amount of articles in biotechnology field and it also includes affiliation for each co-author, which is not the case for most of other scientific article databases. The affiliations of the authors are of a great importance for this research, because the objective is to focus exclusively on Canadian biotechnology innovation. Only the articles and patents with at least one author or inventor with a Canadian affiliation were extracted and included in the databases, on which this research is based. The data covers the publications for the period from 1966 to 2005 which includes a total of 100652 articles that are related to biotechnology and are written by a total of 94484 scholars. They have registered 4893 patents in the period between 1971 and 2006. This amount of data results in very large networks which could be analyzed only by certain types of software. We utilized the Software named Pajek which is specifically developed for analysis of large networks. It is capable of analysis and visualization of networks with millions of nodes.

2.2. Methodology

The purpose of this study is to explore the network of scientists' co-authorships and to investigate the relationship between the structure of this network and three indicators: innovation productivity, innovation quality and research productivity. These three measures have been quantified by the number of patents, number of patent claims, and the numbers of publications, respectively.

To reach the mentioned goals, the study has been conducted in two general phases. During the first phase (which is explained in this section), the collaboration network of scientists has been constructed and social network analysis has been performed. Many network indicators have been measured and collected as input data for the second phase.

In the second phase, the association of the measured network properties with research productivity, innovativeness as well as innovation quality is examined. This phase encompasses a quantitative method using statistical analysis based on the data obtained from the previous phase. The detailed procedure of the second phase is described in the following section of this chapter.

2.2.1. Construction of scientists' co-authorship networks

In this section, the procedure of building the co-authorship networks of biotechnology scientists is explained and different properties of the network structure are examined. Being aware of the advantages and disadvantages of the structure of co-authorship networks, social network specialists and the owners and managers will be able to set up appropriate policies on the relationships and collaborations of their scholars and inventors to achieve higher innovative and research productivity.

Therefore, the networks of co-authorships are constructed based on the scholarly articles that have been co-authored by individual scientists. A simple definition of a co-authored article is presented by Melin and Persson (1996): co-authorship happens when a scholarly document has two or more authors. These co-authorships form connections between individuals. The outcome of all of these connections is a network in which knowledge is exchanged and the wide spread of information leads to the enhancement of innovativeness and research productivity. It is logical to assume that in the course of the co-authorship procedure, information trade happens among the scholars to a great extent (He 2009). The resulting network is called collaboration network and is usually represented by a corresponding graph. The nodes in these graphs represent the individual scientists, and when two authors collaborate on writing a publication, a link representing their co-authorship will connect them in the network as well.

When the collaboration ties are created among scientists, it will lead to knowledge exchange among them and to the scientific and innovative output. It should be noted that these links will last for different periods of time between various individuals. It has been stated that such relationships usually live more than one year (Schilling and Phelps 2007). Given that hardly ever the termination date of collaboration ties are recorded, we cannot be sure about the duration of existence of ties; for that reason, we should make an assumption as to the period that the links persist (Schilling and Phelps 2007).

In order to do that different approaches have been taken into account. Mostly, this value has been assumed to be either three years (by for example McFadyen et al. 2007, Schilling and Phelps 2007) or five years (by for example Stuart 2000, Baum et al. 2003, Fleming et al. 2007, He 2009). In the current study, the five-year approach has been taken, because this assumption has been widely considered by the previous researchers. The next section discusses the construction of the networks based on the assumption of the five-year life of each link.

2.2.2. Two-mode and one-mode networks

The database of articles contains the list of articles and their authors and some other information (the year of publication, article ID, article abstract, author affiliations, etc.). In the first step, the data have been cleaned and duplicates have been removed.

In order to obtain the data corresponding to the five-year periods, some queries have been run in Microsoft Access to extract the publications in each five-year window starting from 1966 to 2005. The first window covers the co-authorships from 1966 to 1970, second from 1967 to 1971, and the last one from 2001 to 2005. This approach resulted in a total of 36 networks which are all undirected networks, meaning that the connecting lines among individuals are simple lines and not arrows. This is because we are only interested in the co-authorships of articles and no other factors (like the person who proposed the publication idea first). Besides, we want to test the collaborations' effect and the strength of the relationships is beyond the scope of this project, if there are multiple lines between two nodes (meaning that two scientists co-authored more than one paper in the five-year period), it is considered as a single link.

Next, the data should be prepared to a special format so that they can serve as input into social network analysis software to be further analyzed. In this study the Pajek software has been selected for the network analysis. The main reason is that Pajek is capable of performing many calculations on even very large networks with millions of nodes. The input data for Pajek had to be in a text file. Before importing the data into a text file, articles and authors needed to get a proper ID in order for Pajek to recognize them. First, appropriate IDs have been allocated to the authors. Their IDs start from 1 and continue up to the number of the last author. Then the IDs have been assigned to the articles. The ID for the first article in each network equals the ID of the last author plus one, and increases in ascending order for the next articles.

After the procedure of assigning IDs, the data was imported into a text file to be ready for the Pajek software. According to the data, Pajek will create a network in which the nodes are representatives of authors and articles; and articles are connected to their authors. This way, the network demonstrates the co-authorships (by inter-connecting the individuals who co-authored an article). This kind of network is called a two-mode network.

Two-mode networks consist of two sets of nodes which are called actors and events (De Nooy et al. 2005). In this kind of network, affiliations connect the two sets of nodes to each other. For example, in our network, the authors are connected to the articles they wrote. Therefore, only nodes from different sets are connected to each other and there is no connection between nodes of the same set (in our network, for instance, the articles are not directly connected to each other).

Recall that in this study we need to evaluate the properties of the network which is constructed based on individuals' collaborations. But, in the two mode network we have nodes of both individuals and articles. Therefore, it would be so difficult to interpret such a network, since every parameter measured on this network will incorporate the articles' nodes as well. Consequently, we need to extract a network containing only nodes of authors from this network.

To do this, we converted the current network to two separate ones, one considering only actors (authors in this study), and the other one containing events (articles in this study). This kind of networks is called one-mode network (De Nooy et al. 2005). By doing this conversion the nature of connections will not change. For instance, in the authors' network, individuals who co-authored a scholarly document will be connected to each other; hence, we are now able to concentrate on the scientists' network which is of our interest in this study. Examples of both two-mode and one-mode networks and the conversion from two-mode to one-mode are illustrated in Figure 3 and Figure 4. Since our networks are very large, it is impossible to show their graph in regular size papers. However, examples of two-mode and one-mode networks of biotechnology scientists in early periods of this study are shown in appendix A.

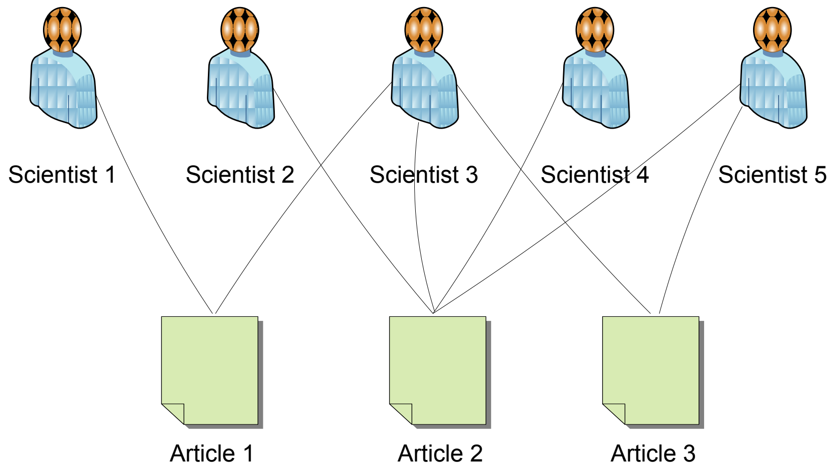

Figure 3: Two-mode affiliation network of article co-authorships

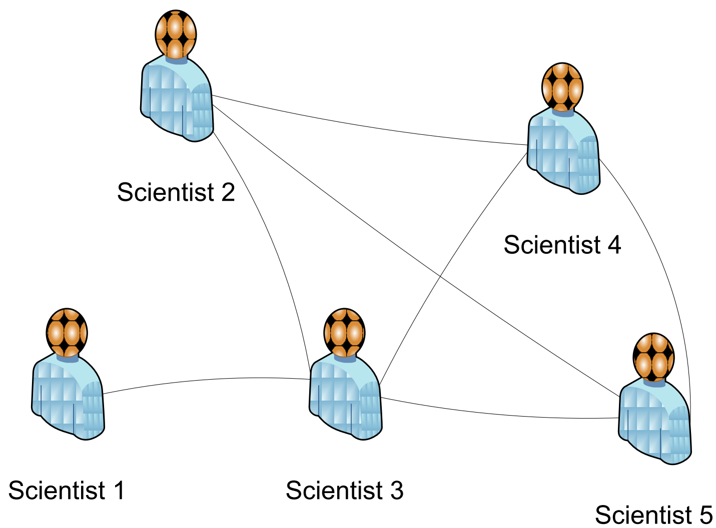

Figure 4: One-mode network of scientists

The next step is to calculate various properties of the constructed networks to achieve the defined objectives of this study that will be described in detail in the next chapter.

Chapter 3: Analysis of data

After the networks were constructed, their structure was analyzed with the help of Pajek. The structural network properties that needed to be assessed in order to answer our research questions were measured and recorded. It should be mentioned that in order to analyze the effect of network structure on the network productivity, first we need to quantify these concepts. To measure the productivity of the network, we need to define the necessary performance measures which are presented below as the dependent variables. In order to quantify the structure of the networks, we need to define indexes that represent various properties of network structure; these indicators are presented as independent variables. Therefore, we aim to investigate the relationship between dependent and independent variables.

In this chapter, first the necessary variables are defined and explained. Each of the variables described below has been computed with the help of Pajek for each of the networks built on the five-year moving windows. In the second part of this chapter, various statistical analyses which have been performed to the data to address the goals of this study will be described.

3.1. Variables

3.1.1. Dependent Variables

As mentioned before, the aim of this study is to evaluate the impact of structural properties of scientists' collaboration network on three different measures (dependent variables). These measures are presented and described below:

Innovative performance of scholars: The first dependant variable is the individual innovative performance of scholars in the field of biotechnology in Canada. The patenting activity of the individuals is considered as a proxy of this variable, i.e. the number of patents that the inventors have published is a representative indicator showing their innovative performance. The variable P_i will be the indicator of number of patents in year i; where i ranges from 1971 to 2006. The number of patents is considered to be a rational proxy for innovative performance and has been widely used before (Fleming et al. 2007, Schilling and Phelps 2007, He 2009, Jaffe et al. 2000, Ahuja 2000).

Quality of the innovative production: The second dependent variable is the quality of the innovative production of individual inventors. This value has been measured by the number of patent claims of the patents[4]. It was assumed that the greater the number of patents claims, the higher the quality of the patent in terms of its innovation potential. According to Tong and Frame (1994), the number of claims associates positively with the value of the patent and well predicts its future commercial potential. In the innovation

[4] "Patent claims are a series of numbered expressions describing the invention in technical terms and defining the extent of the protection conferred by a patent (the legal scope of the patent). A high number of patent claims is an indication that an innovation is broader and has a greater potential profitability" (Beaudry, C., & Schiffauerova, A. 2011).

research the patent claims have been used previously as an indicator of patent quality. This value is noted as PQ_i and denotes the number of patent claims in year i.

Research productivity of scholars: Third dependent variable is the research productivity of scholars in Canadian biotechnology sector. To measure this variable, the number of the articles published by the scholars has been taken into consideration. According to Toutkoushian et al. (2003), the number of publications is the most common measure of scientists' research productivity. As an example, the following researchers have used this quantity as a measure of research productivity: Centra (1983), Bland and Ruffin (1992), Taylor et al. (1984), Kuzhabekova (2011), and Rumsey-Wairepo (2006). The notation used for this variable in this study is $ARTC_i$, which reflects the average number of articles published by individuals in year i.

3.1.2. Independent Variables

The purpose of this study is to measure the impact of the structural properties of the Canadian biotechnology network of scientific co-authorships on the scientific and innovative performance of the scientists. In this section all the independent variables whose effect on the scientific or innovative performance will be studied are described and discussed in detail.

It should be noted that we examine the effect of each structural network variable on the scientific and innovative performance in the first year which follows the interval in which the network architecture was assessed. For example, we evaluate the impact of the structure of collaboration network of scientists between 1966 and 1970 on the innovative or scientific performance of the individuals in 1971. The reason for this is an assumption that the fruits of the scientists' 5-year collaborative period will be gathered only after this period has finished. It usually takes time to publish an article or to register a patent. This assumption is commonly made by other researchers in similar studies as well (Stuart 2000, Baum et al. 2003, Fleming et al. 2007, He 2009). As a consequence, having a dependent variable calculated on year i, related independent variables are calculated for the networks constructed on the five-year snapshot from year i-5 to i-1. For instance, for the dependent variable on year 1971, corresponding independent variables are calculated for the network of years from 1966 to 1970.

Independent variables are listed and described below:

Connectivity of the Network (ConNet): This variable measures the average degree of the whole network. For each node (scientist) the degree is defined as the number of ties connected to it; in other words, it shows the number of nodes adjacent to the particular node (Wasserman and Faust 1994). When the degree of each node is calculated, the average of all values returns the overall degree of the whole network. Obviously, when the average degree increases, the network will be denser meaning that the nodes have more ties, signifying more co-authorship relations between individuals.

Therefore, the average degree is considered as a proper measure for the structural "cohesion"[5] of the network (De Nooy et al. 2005). Higher values of the average degree imply that there are more collaborations in the network; as a result more knowledge

[5] "Cohesion means that a social network contains many ties. More ties between people yield a tighter structure, which is, presumably, more cohesive." De Nooy et al. (2005)

exchange will take place. We presume that this will affect the innovativeness and research productivity of individuals positively.

Size of the Largest Component (LC): Largest component of a network is a sub-network in which there is no isolated node and all the nodes are directly or indirectly connected to each other. In fact, the largest component is the largest fraction of the network where information exchange takes place. Since the size of the largest component has a numerically large scale, in the present thesis the natural logarithm value is used to compress the data scale to be incorporated in the regression model[6]. Therefore this variable returns a natural log value.

Proportion of the Largest Component (PLC): This variable measures the percentage of nodes that are in the largest component:

$$PLC = \frac{LC}{Total\ number\ of\ nodes\ in\ the\ network}$$

Degree Centralization (DC): the concept of centrality refers to the importance of the network members in the process of information exchange in the network. When an actor is widely involved in the communications with other individuals, we can conclude that this actor plays an important role in the knowledge diffusion in the network. According to Wasserman and Faust (1994), this kind of involvement is called the centrality of the vertex. The centrality of a node could be analyzed from different aspects. One of the centrality measures is degree centrality.

[6] The same approach has been utilized for similar situations by various authors such as Ahuja (2000), Fleming et al. (2007) and He (2009).

Degree centrality of a node measures the number of nodes that are directly connected to this node. Clearly, the more a network is connected to other nodes, the more active it will be in the sense of information transfer and consequently, it will be more central. This value has been reflected in the variable *"connectivity of the network"* defined before.

According to He (2009), when the degree centrality of network nodes vary more, the network will be more centralized. Therefore, we evaluate the centrality of the network by its degree centralization which is calculated by dividing the variation of nodes' degrees by the highest possible variation in a network of the same size:

$$DC = \frac{variation\ in\ degree\ cebtrality\ of\ vertices}{maximum\ variation\ in\ degree\ centrality\ of\ network\ of\ the\ same\ size}$$

Betweenness Centralization (BC): Betweenness centrality takes into consideration the role of intermediary individuals, i.e. the scientists that lie on the paths connecting two nodes (Wasserman and Faust 1994). In other words, this measure evaluates the significance of a person as a connector between two other individuals that can enhance the knowledge exchange between them. Therefore, the betweenness centrality of a node is defined as the proportion of all shortest paths between pairs of other nodes that contain this node (De Nooy et al. 2005). This indicator shows the control of a node over the relations between the other individuals within the network and its impact on the information flow among them.

The variation in the betweenness centrality of nodes in a network is measured by betweenness centralization. It is actually calculated by dividing the variation in the betweenness centrality of vertices by the highest possible betweenness centrality variation in a network of the same size (De Nooy et al. 2005):

$$BC = \frac{variation\ in\ \ betweenness\ centrality\ of\ nodes}{maximum\ variation\ in\ betweenness\ centrality\ of\ nodes\ in\ a\ network\ of\ the\ same\ size}$$

Closeness centralization (CloC): Another way to measure the centrality of an individual in the network is to evaluate the distance of each node to all the other nodes in the network. The indicator that covers this concept measures how close a node is to other members of the network and it is called closeness centrality. If a scientist is close to many other scientists in the network, he/she will have an easy access to a vast amount of information and consequently it is assumed that this will improve the productivity of the network. The smaller the distance of a node to all other nodes the higher its closeness centrality, and therefore the easier it can reach the flow of knowledge.

Like degree centralization, we calculate the closeness centralization of network by calculating the variation in closeness centrality of vertices and dividing it by the highest possible variation in closeness centrality in a network of the same size (De Nooy et al. 2005):

$$CloC = \frac{variation\ in\ the\ closeness\ centrality\ of\ nodes}{maximum\ variation\ in\ closeness\ centrality\ of\ nodes\ in\ a\ network\ of\ the\ same\ size}$$

It should be noted that when a network is disconnected, i.e. there are separate components in the network, the distance between nodes of the disconnected components cannot be calculated. To resolve this issue in this study, the closeness centralizations have been computed only for the largest component of each network. This could be easily justified, because the largest components of the most of the networks (especially the ones for the later intervals) cover a large proportion of the entire network (close to the 90% of the network).

Clustering Coefficient (CC): As defined in chapter two, clustering coefficient measures the level of clustering in the network. This index evaluates the level of tendency of the nodes to cluster together. Watts and Strogatz (1998) introduced a method to measure the local clustering coefficient of each node within a network, which is defined as:

$$CC_i = \frac{number\ of\ triangles\ connected\ to\ node\ i}{number\ of\ triples\ centered\ on\ node\ i}$$

$$CC_i = \frac{2z_i}{n_i(n_i - 1)}$$

In this equation n_i is the number of direct neighbor nodes of node i and therefore the term $\frac{n_i(n_i-1)}{2}$ denotes the total number of possible links between node i's neighbors. Z_i represents the total number of existing links between the n direct neighbors of the node i (Clements 2008).

For each node, this index measures the proportion of connections between the neighbor nodes of this node over all possible links that could exist among these neighbors. This variable returns a value between 0 and 1 and it is zero for the nodes that have 0 or 1 neighbor. The average of the local clustering coefficients of all the nodes denotes the overall clustering coefficient of the entire network:

$$CC = \frac{1}{n}\sum_i CC_i$$

There is another approach to compute the clustering coefficient of a network. In this method, unlike the previous one, the global clustering of the network is directly measured. To determine this index, the proportion of triangles in the network over the number of connected triples is calculated. Connected triples are sets of three nodes in which there are at least two connecting links (for example among nodes i, j and k, i is connected to j and k, but j and k are not necessarily connected) and triangles consist of three nodes all connected to the other two nodes (Schilling and Phelps 2007).

Therefore, for each of the networks, overall clustering coefficient could be obtained from the following formula (Newman et al. 2002):

$$Clustering\ coefficient = \frac{3 \times (number\ of\ triangles\ in\ the\ network)}{number\ of\ connected\ triples\ in\ the\ network}$$

The proportion is multiplied by 3 in the numerator to keep the result in the range between zero and one. This is because each triangle includes 3 triples.

In this study the first method of calculating the clustering coefficient has been utilized. It should be noted that since the purpose of measuring the clustering of a network is to evaluate the small-world characteristic of it, and due to the restrictions in determining this index which will be explained later, the clustering coefficient has been measured over the largest connected component of each network.

Shortest path length (PL): The concept of the shortest path is explained in chapter 2. The shortest path between two nodes is the lowest number of nodes that should be traversed to reach from one node to the other. This value is also called the geodesic (De Nooy et al. 2005). The average of path lengths between pairs of vertices is the overall shortest path of the entire network. The shortest path between the nodes returns the distance between them and supposedly, when the distance between two individuals in a network is shorter, the information can flow easier between them. It is assumed that this will result in a more intensive collaboration and a higher performance.

Similarly as with closeness centralization, there is a limitation on calculating the average shortest path of a network. Most of the social networks contain isolated components (Fleming et al. 2007) and the distance between nodes of separate components cannot be calculated. Therefore, like in similar studies (Fleming et al. 2007, Uzzi and Spiro 2006, He 2009, Newman 2000), the shortest path length is measured over the largest connected component of each network.

Another important point is that the increase in the size of the network greatly affects the shortest path length of a network. This is because when the number of nodes increases, large number of links should be created in the network to keep the distance between nodes short. Otherwise, it's possible that the shortest path of network with lots of nodes looks very high comparing with a small network in which nodes are poorly connected (Fleming et al. 2007). To account for the change in the number of nodes in the network, we normalized this variable by dividing the average shortest path of each network by the theoretical path length of a fully connected graph of the same size and average degree.

The theoretical path lengths have been calculated by the approximation method of Fleming et al. (2007). In this method the path length of a regular graph is considered as $\frac{N}{2z}$ if $N > 2z$ and as $2 - (\frac{z}{N-1})$ if $N \leq 2z$ (*N* is the size of our largest component and *z* is its average degree).

Small-World (SW): As defined in chapter 2, the small-world measure is calculated by dividing the clustering coefficient by the average shortest path length of the network. In some studies this ratio is stated as the product of clustering coefficient and inverse path length:

$$SW = \frac{CC}{PL} = CC \times inverse\ path\ length$$

Again, there is a limitation on calculating the small-world measure, since the small-world ratio cannot be defined on a disconnected graph. In this study, we follow the method used by most of the researchers in this area (Fleming et al. 2007, Newman et al. 2001, Uzzi and Spiro 2006, Kogut and Walker 2001, Baum et al. 2003) who consider only the largest connected component of the network for their analyses. The concept of the largest component has been explained before.

3.1.3. Control variable

Network size (Ln_Scts): In order to account for other factors that can affect our dependent variables, we control for the size of the network. As the new scientists join the network of biotechnology in Canada, there will be more chance for collaborations and as a result, more potential opportunity of knowledge exchange. This will clearly have an impact on the overall scientific and innovative productivity. In order to account for these effects the size of the network, i.e. number of scientists in the network, will be introduced into the model. The variable *Ln_Scts* takes natural log values. Like for the size of the largest component, since the size of the network has a numerically large scale, the natural logarithm value is used to compress the data scale to be incorporated in the regression model.

3.2. Descriptive data analysis

The dataset consist of 94484 authors who have published 100652 articles between 1966 and 2005. Out of these scientists, 5013 cooperated on innovative projects and registered a total of 4893 patents from 1971 to 2006. It should be mentioned that there are two dates associated with each patent in the database. One is the date of application of the patent and the other one shows the date that patent has been granted. In this study the application dates of the patents are employed for the patents that have been granted later. This date has been taken into account because when the application date is available for a patent it shows that the particular innovative activity of the inventors has come to its end; and since we know that these patents will be granted later, we are certain about the innovativeness of the publication.

The summary of the data is presented in Table 1.

Table 1: Data summary

Variable	Observations
Total number of scientists from 1966 o 2005	94484
Total number of patents from 1971 to 2006	4893
Total number of articles from 1971 to 2006	100652
Total number of patent claim from 1971 to 2006	67750

Based on this dataset the values of all independent, dependent and control variables are calculated over the networks and their descriptive statistics are shown in Table 2. This table is taken from SPSS 19.

Table 2: Descriptive statistics of the variables

	Minimum	Maximum	Mean	Std. Deviation
ConNet	1.269	9.341	4.829	2.607
DC	0.003	0.064	0.009	0.012
CloC	0.055	0.709	0.168	0.137
BC	0.000	0.075	0.028	0.017
Ln_Lc	2.079	10.538	8.127	2.600
PLC	0.015	0.897	0.558	0.324
PL	0.003	1.249	0.174	0.364
CC	0.741	0.929	0.787	0.035
SW	1.950	2422.747	996.147	912.539
Ln_Scts	4.533	10.647	9.128	1.521
P	0.000	417.000	135.917	146.371
PQ	0.000	7245.000	1881.944	2160.874
ARTC	69.000	28073.000	11449.194	9407.112

In order to have a better understanding of the collaboration networks over the period under study, in the following section the analytical results of the historical trends for each independent and control variable are presented and discussed separately. The historical trends are extracted based on the original five-year intervals used for the construction of the networks. In the following figures, the values on the historical axis indicate the last year in a five-year interval and the vertical axis represents the values of the corresponding variable in the five-year based network. The descriptive analysis of each variable is presented below:

Network size

The first aspect of the Canadian biotechnology network being observed is its size, i.e. the total number of scientists that are engaged in at least one research activity in the corresponding period of time. Figure 5 shows the graph of Canadian biotechnology network size from 1970 to 2005 based on the count number of scientists.

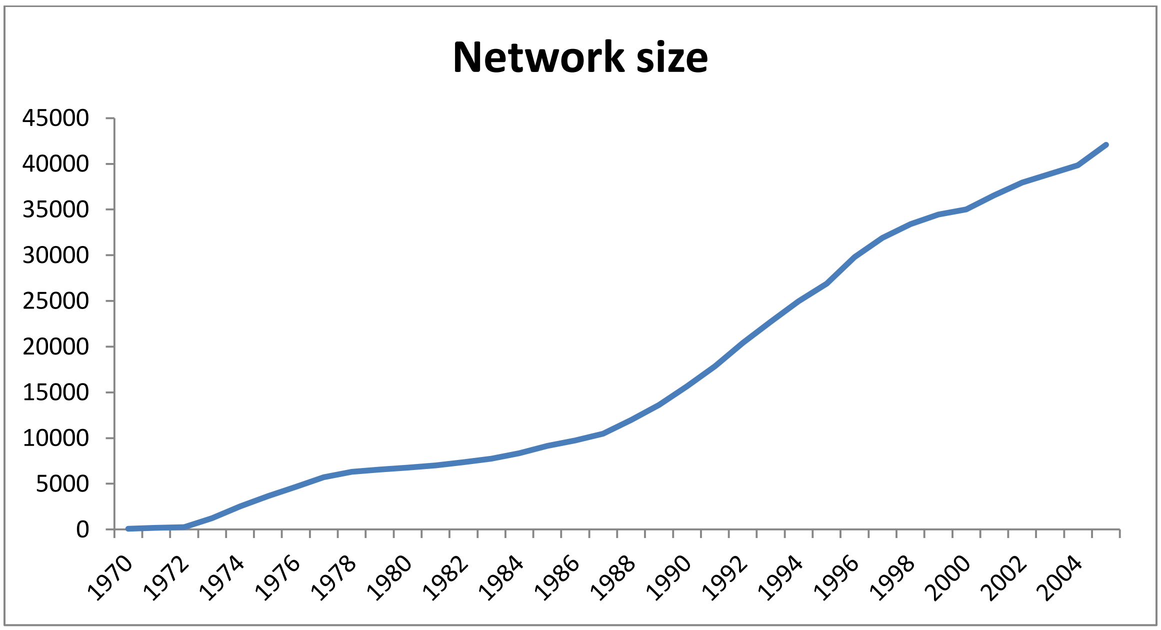

Figure 5: Historical trend of network size from 1970 to 2005

The growth started relatively slowly in 1970s, but as biotechnology became more popular more and more scientists have entered this field in 1980s. The sudden jump in the population growth of scientists in 1980s has been explained as the result of popularity of internet in different research areas (Munn-Venn and Mitchell 2005).

Connectivity of the network (ConNet)

In graph theory, the degree of a vertex is the total number of lines that are directly connected to it. In the Canadian biotechnology network under study, this number denotes the total number of collaborators for each scientist who had at least one article co-authorship during the given period of time. As Wasserman and Faust (1994) discussed, the more the number of co-authors in the network, results in a tighter network in which the knowledge exchange, and consequently innovative productivity is more prone to occur. As Figure 6 shows, the average connectivity of the Canadian biotechnology network increases rapidly over time from almost 1 in 1970 to more than 9 in 2005. Since the size of the network increases and more scientists enter the Canadian biotechnology sector over time, this rise in the network degree is expected.

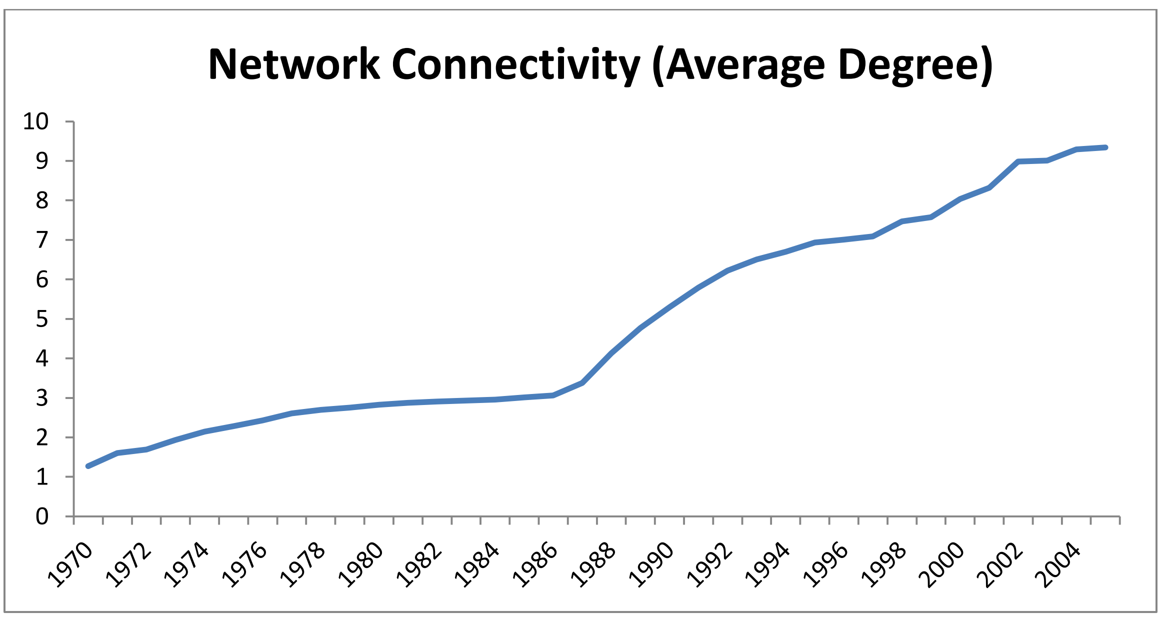

Figure 6: Historical trend of network connectivity from 1970 to 2005

According to the figure, network connectivity increases slowly between 1970 and 1986, and then there is a jump after 1987 which corresponds with the sudden growth of network size in the same year, as illustrated previously.

Closeness Centralization (CloC)

The historical trend of the closeness centralization for the Canadian biotechnology networks is illustrated in Figure 7. In the first years, since the components are sparse and the largest connected component of the network covers a small proportion of all nodes, the closeness centrality will vary greatly among scientists, resulting in a very high closeness centralization of the whole network. As more connections occur in the network due to the growth of the network population, new collaborations are stimulated among existing scientists as well as new ones.

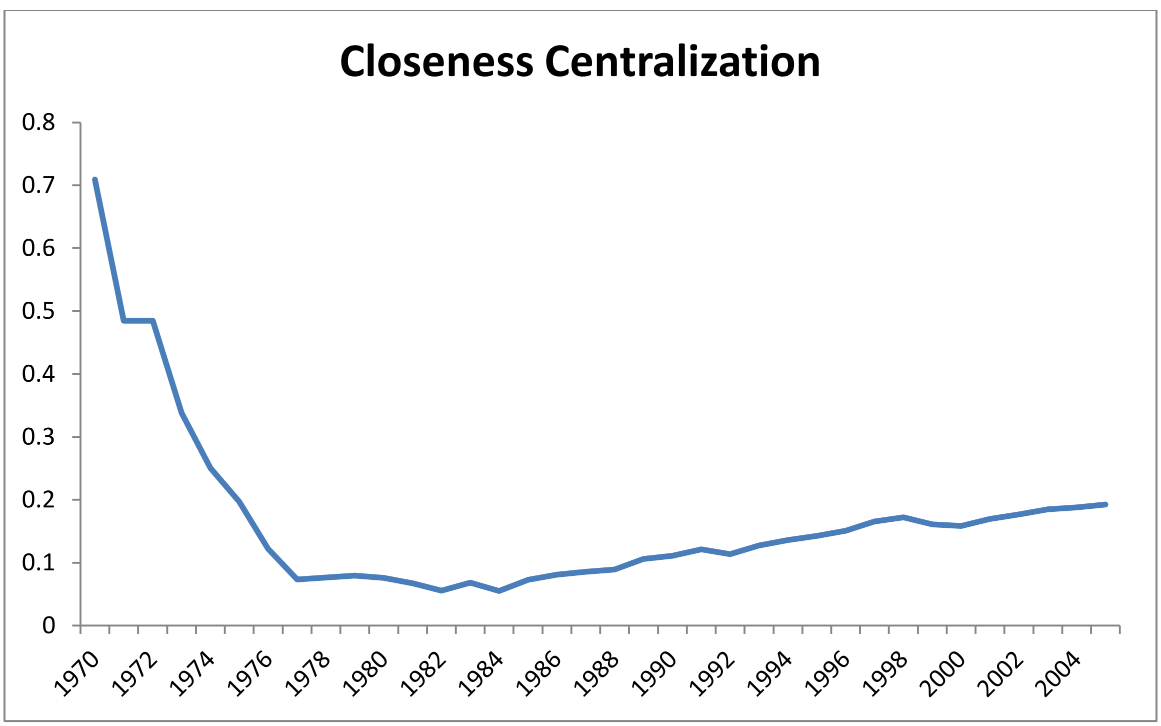

Figure 7: Historical trend of closeness centralization from 1970 to 2005

In the first periods, since the size of networks is small and there are few connecting ties among them, any new alliance between scientists greatly affects the closeness centralization of the network. This decreasing fashion of the closeness centralization continues until 1984. After this period, variation in closeness centrality starts to increase. This increase in the trend may demonstrate the improvement in the number of scientists who work within collaborative groups. Since working in groups could be a result of trust and reciprocal relationships among scientists which is sign of cliquishness (Burt 2001), this increasing trend could be considered as the indication of increase in the cliquishness of the network.

Size and Proportion of the largest connected component (Ln_LC, PLC)

The largest component of a network represents the largest number of connected scientists who have the potential of access to the same knowledge distributed through the network. As the network grows and more connections are created between its nodes from various sub-networks, more small components converge together and build a bigger component.

In figures below, the size of the largest component for each of the intervals, as well as its proportion over the total size of the network are represented respectively. The figures illustrate that the largest component of the Canadian biotechnology network grows gradually in size as the number of scientists increases and their mutual collaborations in the network become more frequent. They also demonstrate that isolated nodes and groups of scientists gradually join the largest component and increase its proportion over the whole network as more research collaborations take place.

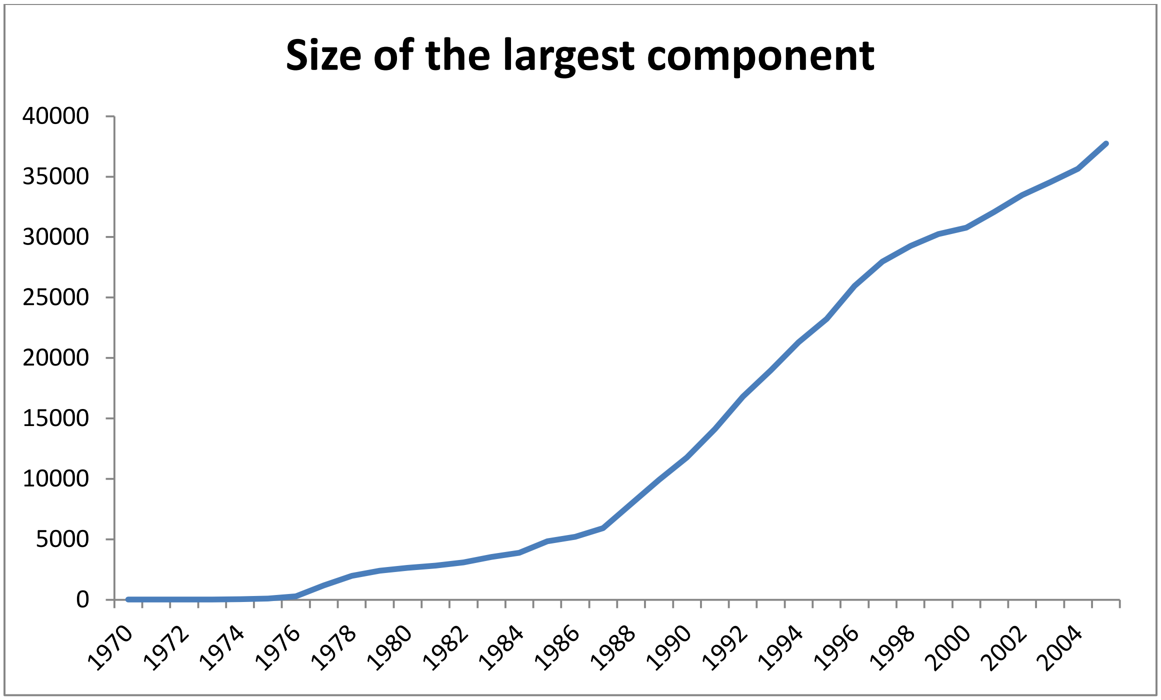

Figure 8: Historical trend of the number of scientists in the largest connected component of Canadian biotechnology network between 1970 and 2005

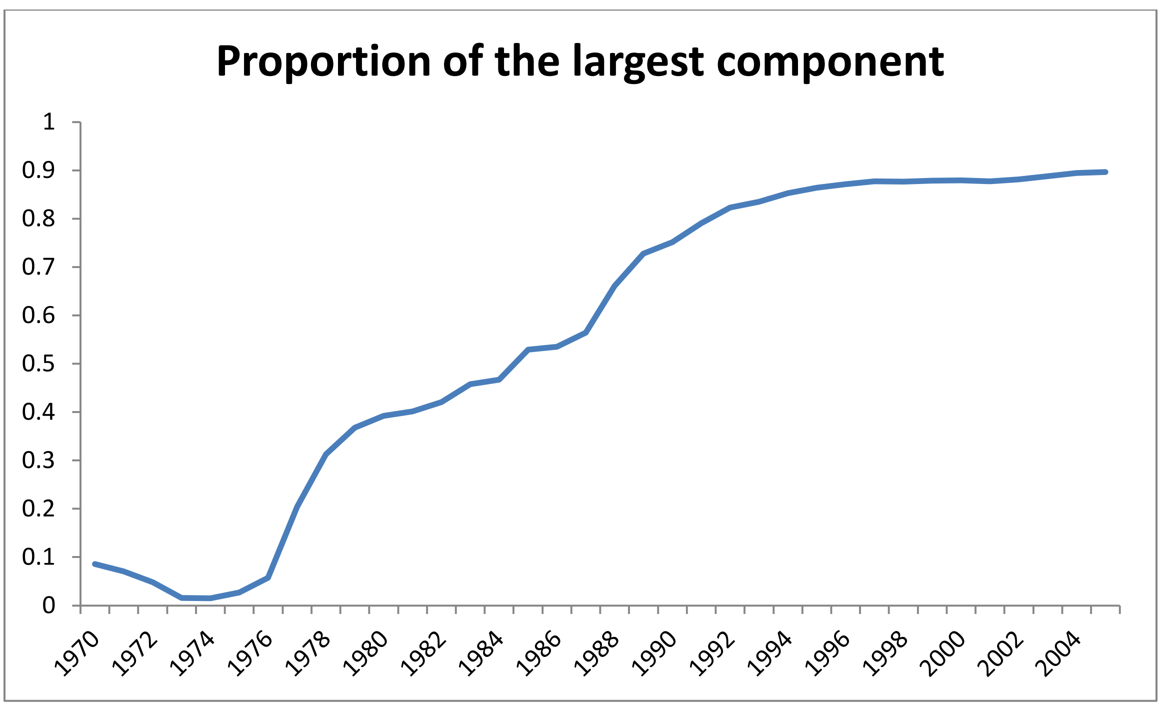

Figure 9: Historical trend of the proportion of the scientists in the largest component in Canadian biotechnology network between 1970 and 2005

Degree Centralization (DC)

The historical trend of degree centralization of the Canadian biotechnology network is illustrated in Figure 10. This variable is an indicator of variation in degree centrality of scientists in the network. According to the figure, this variation drops rapidly in the first years; this could be the result of few numbers of scientists working on biotechnology fields in the first time intervals. Hence entrance of new scientists to the network highly affects the degree centralization of the whole network. After this fast drop, the value of degree centralization remains below 0.01 during the whole studied period. This means that the distribution of the links among scientists is almost the same for the whole network, resulting in a homogenous growth of collaborative opportunity for all scientists in the network.

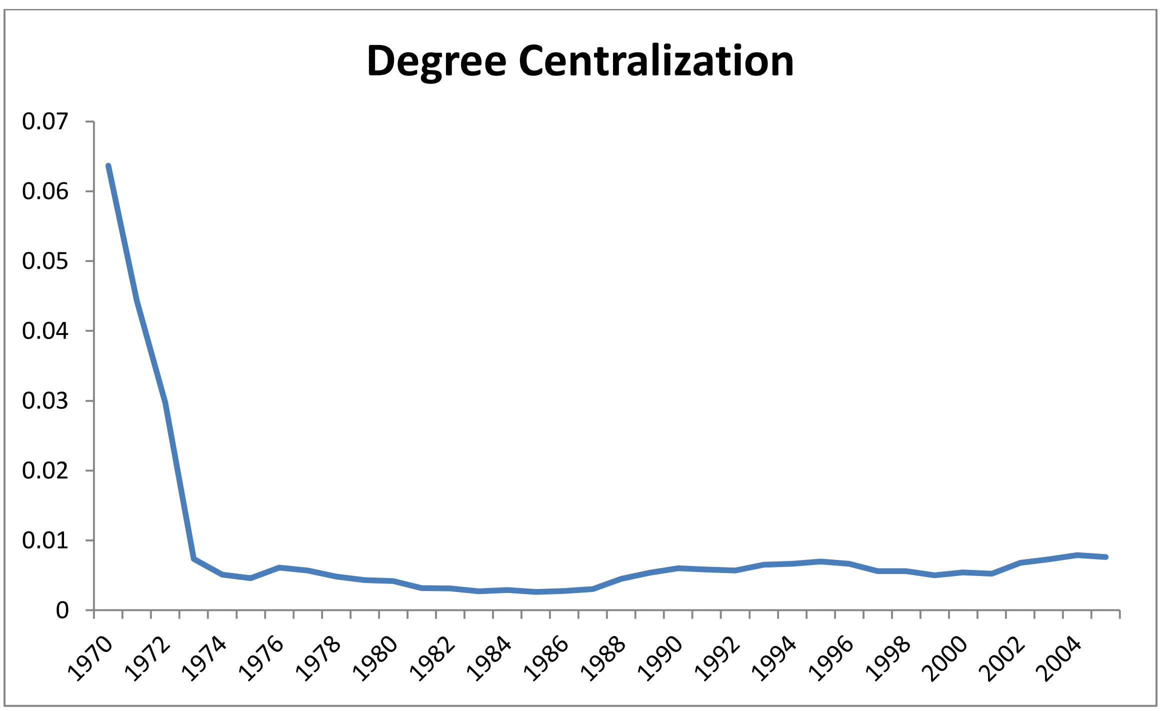

Figure 10: Historical trend of degree centralization of Canadian biotechnology network from 1970 to 2005

Betweenness Centralization (BC)

As mentioned earlier, betweenness centralization of the Canadian biotechnology network describes the variation of betweenness centrality of its vertices (scientists). In other words, it is an indicator of the position of scientists as intermediaries for the flow of knowledge. The historical trend of betweenness centralization of the Canadian biotechnology network is presented in Figure 11. The figure suggests no significant trend over time. There are some significant fluctuations until 1985, but overall betweenness centralization of the network remains almost constant afterwards, implying that the variation of betweenness centrality of scientists is not affected by the growth of the network size. However, this does not tell us how the average betweenness centrality of the individual scientists is affected over time.

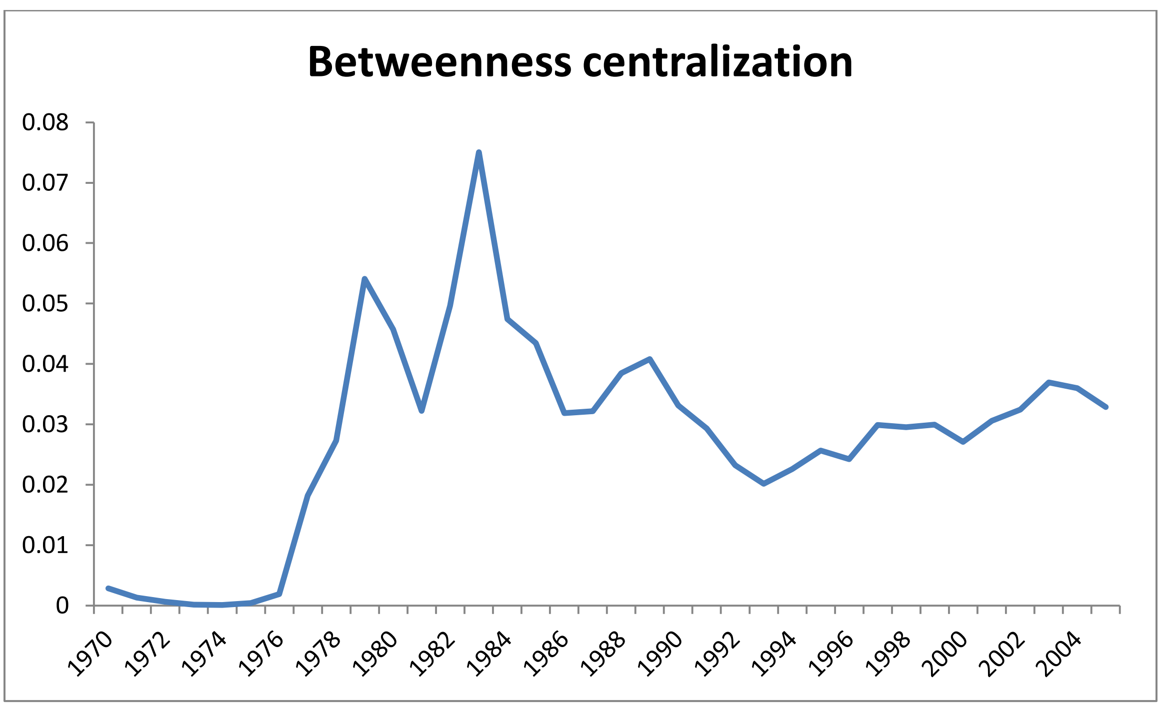

Figure 11: Historical trend of betweenness centralization of Canadian biotechnology network between 1970 and 2005

Shortest path length (PL)

The shortest path between each pair of nodes in the network indicates their distance and thus their potential ability to collaborate with each other. The average shortest path, as it was described before, represents the separation degree of the Canadian biotechnology scientists in general. The historical changes of the average shortest path for Canadian biotechnology network are depicted in Figure 12.

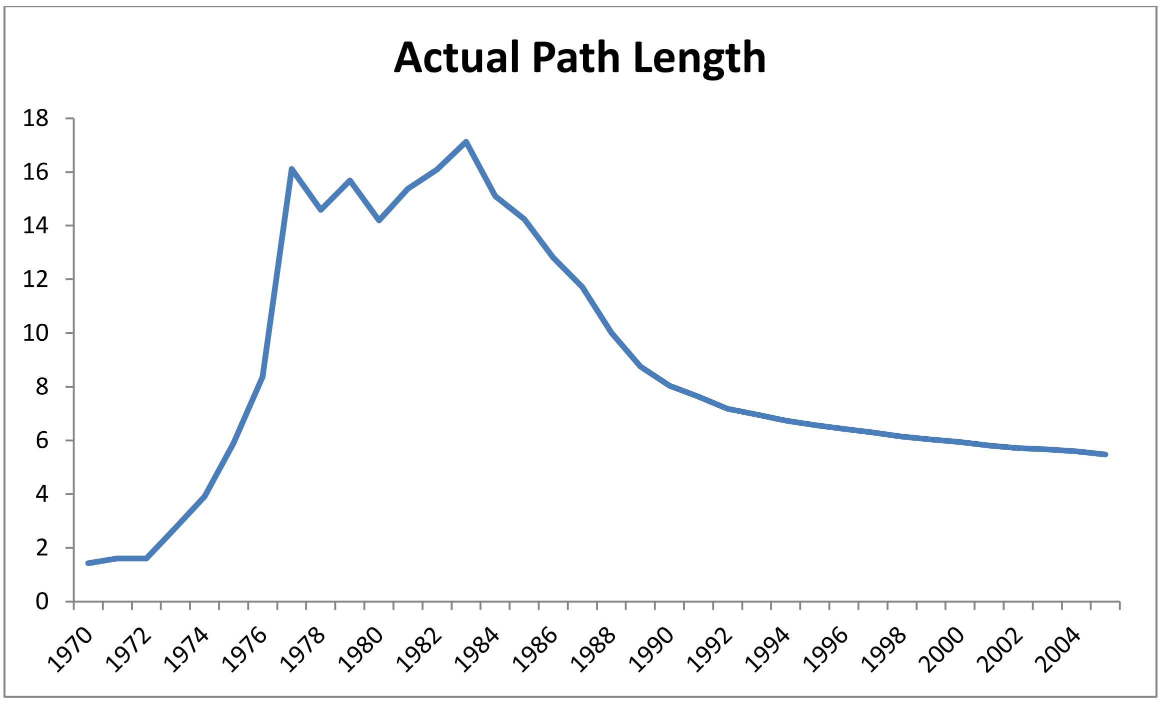

Figure 12: Historical trend of actual values of path length of the Canadian biotechnology network from 1970 to 2005

The graph demonstrates that the path length of the networks increases first. The increase is due to the raise of the number of nodes; evidently, when network size increases, the nodes will become more separated (Albert and Barbasi 2002). The figure also shows that after 1983, the shortest path starts to decrease. The declining trend of this measure is considered to be one of the main indicators of small-world phenomenon in large networks. The shortest path reduction proves that the number of ties between scientists has increased and there are more links created both among the scientists already existing in the networks and between the new individuals entering the networks and the existing ones (Clements 2008).

The results of path length measurements, reveals an important outcome: interestingly, the values of path length converge to 6 in the later periods. This implies that the average distance between individuals is around 6, which is in consistent with the results of Milgram's (1967) who first introduced the small-world structure and based on his empirical study, to reach a person who is unknown to an individual, on average only six intermediates are needed.

Clustering Coefficient (CC)

Clustering coefficient measures the fraction of collaborators of a node who also collaborate with each other. In clustering coefficient formula, the fraction of triangles connected to a node over the number of triples centered to this node is computed. This fraction returns a value between 0 and 1. Networks with higher interconnectivity between their nodes have clustering coefficient closer to 1. For instance, the clustering coefficient of the network built on the 2000-2005 period shows a clustering coefficient of 0.78. This implies that in such network, overall, two individuals who have a common collaborator are 78 times more prone to work together in partnership than those who do not have this mutual partner.

The historical trend of this variable measured for the network under study is depicted in Figure 13. Clustering coefficient drops greatly at first, then increases and remains relatively constant in final periods.

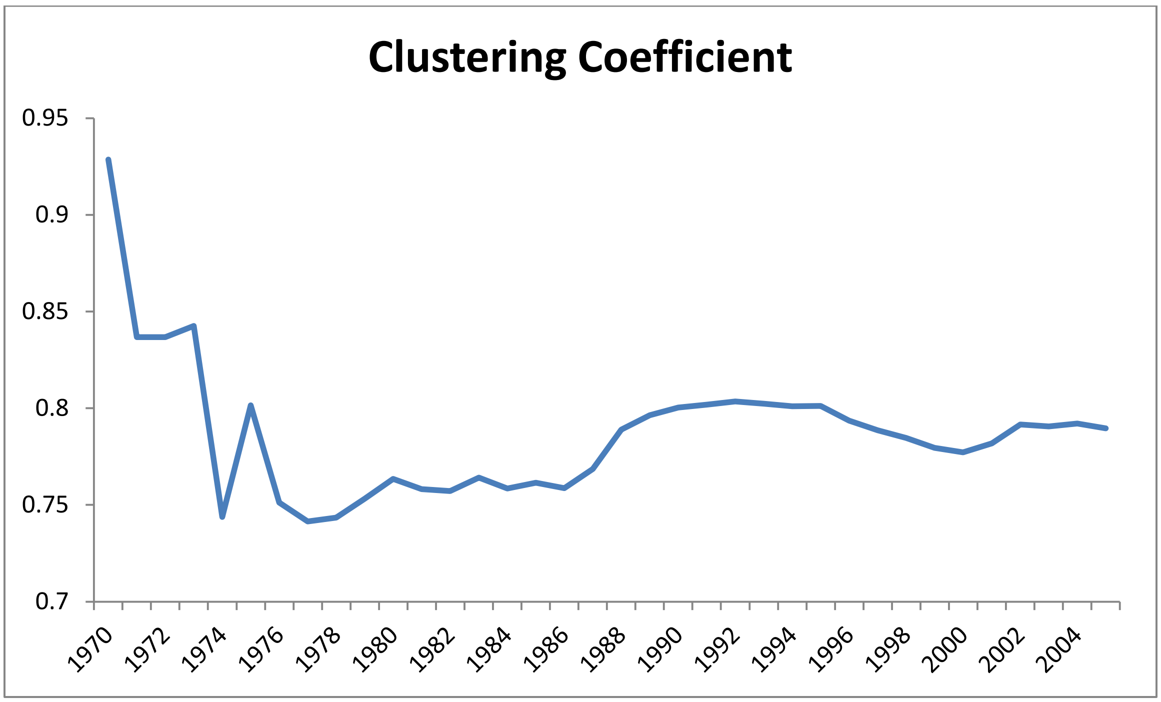

Figure 13: Historical trend of clustering coefficient of the Canadian biotechnology network from 1970 to 2005

Small-World (SW)

The small-world ratio, as mentioned before, is calculated by dividing the clustering coefficient of the network by its average shortest path length. Generally, a network which shows the small-world characteristic demonstrates high clustering as well as short path length between vertices. According to Albert and Barnasi (2002), the small-world networks are often large in size, but despite their size they still exhibit fairly short path lengths and high cliquishness. In order to measure to what extent the structure of our network resembles the structure of a small-world network, we follow the approach of Watts (1999) in which the path length and clustering coefficient are modified first to be used in the small-world equation.

Therefore, in order to incorporate the path length into small-world ratio equation, Watts (1999) presented an approach which has been frequently followed by many other researchers. In this approach the average path length of each network is divided by the average path length of a random network of the same size (n) and the mean degree of the network (z). However, since this index cannot be precisely calculated for a random graph, Watts (1999) approximated it by the following formula:

$$PL_{random} = \frac{Ln(n)}{Ln(z)}$$

Therefore, the path length of a network is reflected in small-world ratio in the form of the actual value of the path length divided by the approximated path length of the corresponding random graph.

The same method is utilized for the clustering coefficient. The clustering coefficient of each network is divided by the one approximated for a random graph of similar size and average degree. The approximation presented by Watts (1999) is as follows:

$$CC_{random} = \frac{z}{n}$$

In order to be a small-world, a network should exhibit a path length relatively equal to the path length of the random graph and also a clustering coefficient much greater than that of the random network (Kogut and Walker 2001). In other words, supposing that the actual clustering coefficient and path length of the network are respectively shown by CC_a and PL_a, and the corresponding values for the random network are CC_r and PL_r, a network is small-world when $CC_a > CC_r$ and $PL_a \sim PL_r$.

The historical trends of the path length and clustering coefficient ratios divided by their values of the corresponding random graphs are illustrated in Figure 14 and Figure 15 respectively. As the graphs show, overall, path length ratio becomes closer to the path length of the random graph (the curve goes to 1) and the clustering coefficient ratio increases with a great rate so that it becomes much larger than that of the random network. Therefore, we can expect that the structure of the collaboration network of Canadian biotechnology resembles the small-world network structure. However, in order to determine this we need to calculate and assess the values of small-world ratio.

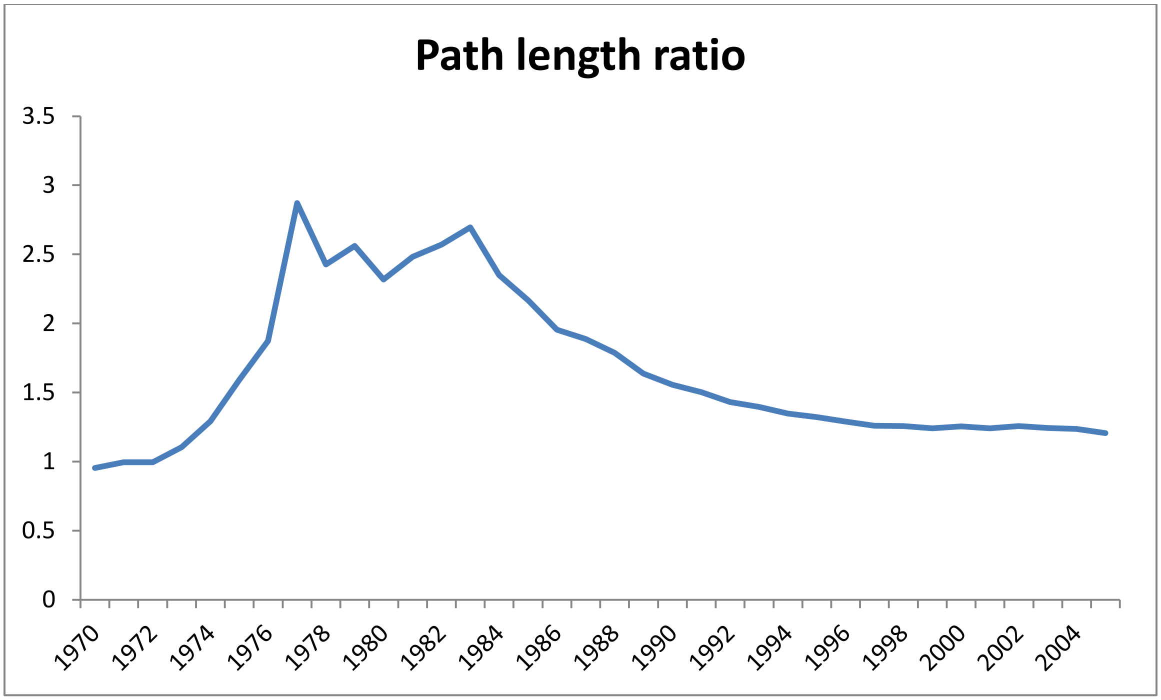

Figure 14: Path length ratio of Canadian biotechnology number (based on the Watts method)

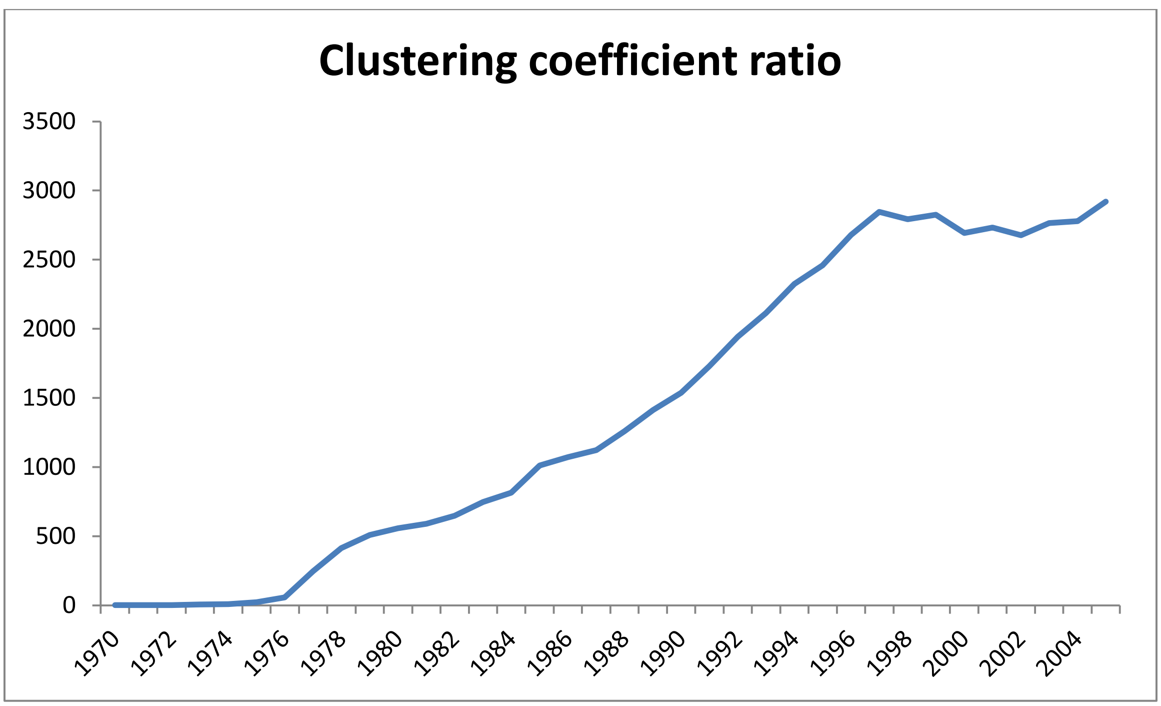

Figure 15: Clustering coefficient ratio of Canadian biotechnology number (based on the Watts method)

In order to gain the value of small-world characteristic for each network we follow the method employed in several previous studies (Davis et al. 2003, Kogut and Walker 2001, Baum et al. 2003), which uses the following ratio:

$$SW = \frac{\frac{CC_a}{CC_r}}{\frac{PL_a}{PL_r}}$$

The results of the calculations of this formula are presented in Table 3. In this table column 1 shows the five-year period on which the network is constructed. Column 2 gives the number of nodes in the largest connected component of the networks (n). Columns 3 and 4 represent the values of path length for actual and random networks. Columns 5 and 6 denote the values of clustering coefficient for actual and random graphs; and the last column gives the values of small-world ratio.

There is no critical index for the small-world measure. Besides, it is implied that when the size of the network increases, the critical value for the small-world should increase (Baum et al. 2003). Therefore, the common procedure (Albert and Barbasi 2002, Davis et al. 2003, Kogut and Walker 2001, Baum et al. 2003) to find out whether the networks exhibit small-world properties or not consists in comparing their small-world values to the networks previously studied in the literature. The list of corresponding values of some of the previously identified small-world networks are summarized in Table 4. The values are taken from the lists gathered by Kogut and Walker (2001) and Albert and Barbasi (2002). By comparing the SW values measured on the networks of this study with the values measured on the networks of similar sizes in the prior research, it can be concluded that the Canadian biotechnology network vividly represents the small-world properties with respect to its high *SW* values.

Table 3: Small-World characteristics for the Canadian biotechnology networks

Time intervals	Number of nodes	Path Length		Clustering Coefficient		*SW*
		PL_a	PL_r	CC_a	CC_r	
1966 - 1970	8	1.43	1.50	0.929	0.5000	1.95
1967 - 1971	12	1.61	1.61	0.837	0.3889	2.16
1968 - 1972	12	1.61	1.61	0.837	0.3889	2.16
1969 - 1973	19	2.75	2.49	0.843	0.1717	4.44
1970 - 1974	38	3.92	3.03	0.744	0.0873	6.60
1971 - 1975	98	5.89	3.70	0.802	0.0352	14.33
1972 - 1976	267	8.37	4.47	0.751	0.0131	30.69
1973 - 1977	1169	16.11	5.61	0.741	0.0030	85.66
1974 - 1978	1970	14.59	6.01	0.743	0.0018	170.86
1975 - 1979	2407	15.68	6.12	0.753	0.0015	198.47
1976 - 1980	2645	14.20	6.13	0.763	0.0014	240.92
1977 - 1981	2808	15.38	6.19	0.758	0.0013	237.94
1978 - 1982	3089	16.09	6.26	0.757	0.0012	252.19
1979 - 1983	3541	17.13	6.35	0.764	0.0010	277.34
1980 - 1984	3891	15.10	6.42	0.758	0.0009	346.63
1981 - 1985	4831	14.25	6.57	0.762	0.0008	466.88
1982 - 1986	5206	12.82	6.56	0.759	0.0007	547.92
1983 - 1987	5915	11.72	6.21	0.769	0.0007	595.59
1984 - 1988	7922	10.02	5.61	0.789	0.0006	704.93
1985 - 1989	9909	8.75	5.35	0.796	0.0006	862.24
1986 - 1990	11760	8.04	5.17	0.800	0.0005	988.64
1987 - 1991	14131	7.63	5.08	0.802	0.0005	1150.79
1988 - 1992	16804	7.17	5.02	0.803	0.0004	1358.08
1989 - 1993	18980	6.97	4.99	0.802	0.0004	1514.06
1990 - 1994	21297	6.73	5.00	0.801	0.0003	1726.10
1991 - 1995	23229	6.57	4.97	0.801	0.0003	1860.51
1992 - 1996	25954	6.43	4.98	0.794	0.0003	2077.63
1993 - 1997	27980	6.29	5.00	0.789	0.0003	2260.65
1994 - 1998	29266	6.14	4.88	0.785	0.0003	2222.68
1995 - 1999	30250	6.03	4.86	0.779	0.0003	2275.84
1996 - 2000	30772	5.93	4.73	0.777	0.0003	2146.59
1997 - 2001	32087	5.81	4.68	0.782	0.0003	2202.37
1998 - 2002	33466	5.71	4.55	0.792	0.0003	2130.35
1999 - 2003	34541	5.67	4.56	0.791	0.0003	2226.03
2000 - 2004	35640	5.59	4.52	0.792	0.0003	2248.34
2001 - 2005	37730	5.47	4.54	0.790	0.0003	2422.75

Table 4: A comparison of previously studied small-world networks

Network	*CCa/CCr*	PLa/PL_r	*SW*	Network size	Reference
Ythan estuary food web	3.67	1.08	3.4	134	Montoya and Sole 2000
E.coli substrate graph	12.31	0.96	12.83	282	Wagner and Fell 2000
E. coli, reaction graph	6.55	1.32	4.96	315	Wagner and Fell 2000
Power grid	16	1.51	10.6	4941	Watts and Strogatz 1998
NCSTRL co-authorship	1653.34	1.16	1425.3	11994	Newman 2001
Words, synonyms	1166.7	1.18	988.73	22311	Yook et al. 2001
LANL co-authorship	2388.9	1.23	1942.2	52909	Newman 2001
SPIRES co-authorship	242	1.89	128.05	56627	Newman 2001
Math co-authorship	10925.93	1.16	9418.91	70975	Barabasi et al. 2001
MEDLINE co-authorship	6000	0.94	6382.98	1520251	Newman 2001

As some examples of the comparisons of our result to the ones of other researchers, we can observe a small-world measure of 12.83 for the network of E.coli substrate graph studied by Wagner and Fell (2000), whereas the network of our study with similar size shows the value of 30.69 for this variable; or the network of SPIRES co-authorship analyzed by Newman (2001) demonstrates a SW value of 1942.2, whereas the network of similar size in our study has the value of 2422.75 for this variable. These comparisons and the increasing trend of the SW values for our networks remain no doubt that they have small-world characteristics (Albert and Barbasi 2002, Davis et al. 2003, Kogut and Walker 2001, Baum et al. 2003).

Chapter 4: Statistical Analysis

The purpose of this study is to explore the empirical data gained from Canadian biotechnology network in order to analyze the relationship between research productivity and structural properties of the network. For this purpose the regression technique has been utilized to measure the significance of various structural factors of the network in overall network productivity. First, the correlation analysis between variables is presented to show the association between variables. Then the regression results are presented for each of the dependent variables i.e. the number of articles, number of patents and number of patent claims.

4.1. Correlation Analysis

In order to determine the association between dependent and independent variables, a correlation analysis is performed. The results of correlations calculated with the help of SPSS 19 are illustrated in the table of correlations below.

The corresponding correlations for the dependent variables are highlighted in Table 5. As for the number of patents (*P*), the table demonstrates positive correlations with network connectivity (ConNet), size of the largest component (Ln_LC), proportion of the largest component (PLC), and small-world (SW) characteristic of the network. It is also shown that there is a negative correlation between number of patents and path length (PL). However, the table does not display any significant correlation between the number of patents and the rest of independent variables i.e. degree centralization (DC), closeness centralization (CloC), betweenness centralization (BC) and clustering coefficient (CC).

Table 5: Correlation analysis of Canadian biotechnology network

		ConNet	DC	CloC	BC	Ln_Lc	PLC	PL	CC	SW	Ln_Scts	P	PQ	ARTC
ConNet	Correlation	1	-0.297	-0.21	0.227	.812*	.907*	-.554*	-0.012	.981*	.810*	.931*	.918*	.971*
DC	Correlation	-0.297	1	.919*	-.486*	-.622*	-.388	.760*	.804*	-0.244	-.762*	-0.192	-0.176	-0.285
CloC	Correlation	-0.21	.919*	1	-.646*	-.675*	-.403	.864*	.846*	-0.155	-.728*	-0.086	-0.062	-0.191
BC	Correlation	0.227	-.486*	-.646*	1	.622*	.469*	-.693*	-.469*	0.194	.514*	0.127	0.155	0.247
Ln_Lc	Correlation	.812*	-.622*	-.675*	.622*	1	.938*	-.910*	-.391	.792*	.956*	.698*	.665*	.786*
PLC	Correlation	.907*	-.388	-.403	.469*	.938*	1	-.726*	-0.091	.909*	.865*	.812*	.766*	.863*
PL	Correlation	-.554*	.760*	.864*	-.693*	-.910*	-.726*	1	.607*	-.515*	-.901*	-.430*	-.404	-.538*
CC	Correlation	-0.012	.804*	.846*	-.469*	-.391	-0.091	.607*	1	0.034	-.490*	0.049	0.025	-0.066
SW	Correlation	.981*	-0.244	-0.155	0.194	.792*	.909*	-.515*	0.034	1	.778*	.949*	.921*	.974*
Ln_Scts	Correlation	.810*	-.762*	-.728*	.514*	.956*	.865*	-.901*	-.490*	.778*	1	.701*	.673*	.786*
P	Correlation	.931*	-0.192	-0.086	0.127	.698*	.812*	-.430*	0.049	.949*	.701*	1	.934*	.934*
PQ	Correlation	.918*	-0.176	-0.062	0.155	.665*	.766*	-.404*	0.025	.921*	.673*	.934*	1	.939*
ARTC	Correlation	.971*	-0.285	-0.191	0.247	.786*	.863*	-.538*	-0.066	.974*	.786*	.934*	.939*	1

*. Correlation is significant at the 0.01 level (2-tailed).

Small-world, proportion of the largest connected component and connectivity of the network variables express the highest correlations (more than 0.9). However, since these three indicators also have mutual high correlations with each other, it is not possible to decide which one acts as the main predictor for the total number of patents at this stage. Further regression analysis is required to measure the importance of each indicator and its effect on the growth of network patenting productivity. The negative sign of correlation between path length and the number of patents was expected, as any decrease in separation degree between inventors, results in higher knowledge exchange among them.

The table of correlations displays positive significant correlation between patent quality (PQ) as the dependent variable, and network connectivity (ConNet), size of the largest component (Ln_LC), proportion of the largest component (PLC), and small-world (SW) indicators of the network. Besides, the results show that the path length (PL) of the network has a negative significant correlation on subsequent patent quality of the network. Again, these indicators of the network show significant mutual correlations with each, but to determine their effect and the magnitude of this effect on the performance of innovation network we will need to perform regression analysis.

Finally, the table of correlations shows significant positive correlations between the dependent variable of total number of articles (ARTC) and network connectivity (ConNet), size of the largest component (Ln_LC), proportion of the largest component (PLC) and small-world (SW) indicators of the network. Also, like the previous dependent variables, the path length variable (PL) shows significant negative correlation.

Note that the independent variables with strong correlation to all the dependent variables are very similar in both sign and quantity. This suggests that regardless of the type of the measure of innovativeness, the innovative performance of the network might be affected by the same indicators. In other words, as various network measures vary, the total number of patents and their quality as well as total number of articles change in the same direction. This fact is also statistically confirmed by the very high positive correlations between the three dependent variables.

4.2. Regression Analysis

The multiple regression analysis is done using STATA 11 in order to examine the association between Canadian biotechnology network structure and its productivity in terms of the number of articles, patents and their quality. Beside the correlation analysis which only gives an insight to the relationships between pairs of variables in simple binary term, i.e. whether relationship exists or does not exist, the multiple regression analysis will also determine the power of each independent indicator mathematically.

In this section three separate regression models are developed (one for each of the dependent variables) and the results are presented below.

According to Hausman et al (1984), in order to provide a natural baseline for a count measure, the regression model of choice is a Poisson model. Since our three dependent variables, i.e. the total number of articles, patents and patent claims (patent quality), are count measures, the best matching regression model would be Poisson. However, the primary assumption for Poisson model is that it accepts no heterogeneity in the data, which means that variance and mean of the sample should be the same (Coleman 1981).

In reality, however, it is rare to satisfy the Poisson assumption on the actual distribution of a natural phenomenon, because most of the time an overdispersion or underdispersion is detected in the sample data. This causes the Poisson model to underestimate or overestimate the standard errors and thus results in misleading estimates for the statistical significance of variables (Coleman 1981). Hausman et al. (1984) suggest correcting the estimates by using negative binomial regression models instead of Poisson in order to obtain robust standard errors.

According to the descriptive statistics of the data Table 2 all of our three dependent variables show overdispersion, i.e. their unconditional variances are larger than their sample means. Therefore, a likelihood ratio test is conducted for each of the variables to confirm the overdispersion issue using STATA 11. The test outcomes are reported in Appendix B. According to the likelihood ratio test we observe that the overdispersion coefficient (α) for total number of patents and total number of articles report a small value close to zero, by which we accept the null hypothesis that Poisson is a better estimation model in their cases. However, the overdispersion coefficient (α) value for patent quality is relatively significant, which suggests that the negative binomial regression model is a better estimator in this case.

The regression results for each of the three dependent variables are then extracted according to their best matching model (Poisson or negative binomial) using STATA 11 and are presented below.

4.2.1. Regression results for *Total number of articles* model

The observation of the regression coefficients for the impact of the Canadian biotechnology network structural properties on the network's research performance in terms of number of articles published is presented in Table 6. Since the p values reported for all the independent variables (except for the number of scientists in largest component, *Ln_LC*) are less than 0.01, they are considered as significant predictors in the knowledge productivity of the following year.

Table 6: Poisson regression results for *total number of articles* model

ARTC	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ConNet	.0948536	.0122168	7.76	0.000	.0709092	.1187981
DC	-41.65348	3.142961	-13.25	0.000	-47.81357	-35.49339
CloC	1.552135	.2546849	6.09	0.000	1.052962	2.051308
BC	-1.061163	.275935	-3.85	0.000	-1.601986	-.5203404
Ln_LC	-.0825834	.038537	-2.14	0.032	-.1581145	-.0070523
PLC	.5802114	.1674226	3.47	0.001	.2520691	.9083537
PL	-1.857256	.2211598	-8.40	0.000	-2.290721	-1.423791
CC	-4.737143	.3117638	-15.19	0.000	-5.348189	-4.126098
SW	.0003519	.0000321	10.96	0.000	.000289	.0004148
Ln_Scts	.170365	.0704438	2.42	0.016	.0322976	.3084324
_cons	10.87768	.8413502	12.93	0.000	9.228667	12.5267

The first predictor variable, the overall network connectivity of the scientists (ConNet), has a small positive influence in the model of articles. This suggests that the higher number of collaborators per each scientist in the network can lead to the increased research productivity of the network. However, since the value of the corresponding coefficient is small, the effect of this variable is not noticeable.

The table reports a very strong negative influence of the degree centralization (DC) on the productivity. We can conclude that the central structure of the network reduces the overall knowledge spillover among the scientists, resulting in less productivity in the upcoming year.

Considering the closeness centralization (CloC) and betweenness centralization (BC) regressors in the model, we can conclude that the Canadian biotechnology scientists' network takes advantage of the variation in closeness centrality of the individual scientists, but not from the variation in their betweenness centrality. As mentioned before when closeness centralization increases, it might imply that many scientists are working in groups which are the results of trust and their reciprocal relationships (Burt 2001) which could be in favor of knowledge exchange and research productivity of the network, according to the positive sign of its coefficient.

On the other hand, the negative sign of the betweenness centralization coefficient suggests that the more homogeneous the intermediary roles of the individuals are, the better knowledge diffuses among the scientists, which results in higher knowledge productivity.

The positive coefficient of the proportion of the largest component implies that the relative size of the largest component has a significant effect on publishing articles in Canadian biotechnology network. In other words, the integration of disconnected components into one larger component enhances the publishing rate of research articles. As the new entrants join the main component of the network, their chance of absorbing knowledge spillovers significantly improves, which positively affects their research performance.

The results for the small-world measure suggest a negative influence of path length (PL) and clustering coefficient (CC) on the research productivity of the subsequent year. For path length, this result was expected since a decreased average path length among scientists will obviously result in faster and easier exchange of knowledge, which is in accordance with the major conclusions of other researchers (like Watts 1999, Uzzi and Spiro 2005, Schilling and Phelps 2007).

Although the increase in clustering coefficient affects network productivity negatively, the combined effect of path length and clustering coefficient results in a positive small-work measure that improves the network productivity of the following year. However, the very small effect of small-world (0.00035) does not make a strong support for the impact of small-world characteristic on the research performance in innovation networks.

4.2.2. Regression results for the *Total number of patents* model

The results of the regression analysis performed using STATA 11 for the model built on the innovation productivity of individuals (total number of patents) is shown in Table 7. Unlike the results of the articles model, only a few of independent variables demonstrate significant effect based on their *p* values less than 0.01. Among all of the independent variables in this model, connectivity of the network (ConNet), degree centralization (DC) and average path length (PL) seem to impact the patenting of the subsequent year.

Table 7: Poisson regression results for total number of patents model

P	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ConNet	-.6307118	.1313661	-4.80	0.000	-.8881846	-.373239
DC	128.8326	33.03199	3.90	0.000	64.09108	193.5741
CloC	1.592235	2.708528	0.59	0.557	-3.716382	6.900852
BC	-1.364631	3.547213	-0.38	0.700	-8.31704	5.587778
Ln_LC	1.026701	.4424148	2.32	0.020	.1595842	1.893818
PLC	-3.939541	1.849105	-2.13	0.033	-7.563721	-.3153616
PL	8.285182	2.454711	3.38	0.001	3.474037	13.09633
CC	6.48096	3.520791	1.84	0.066	-.4196639	13.38158
SW	-.0007947	.0003594	-2.21	0.027	-.001499	-.0000903
Ln_Scts	4.1847	.7829731	5.34	0.000	2.650101	5.7193
_cons	-44.45991	8.990824	-4.95	0.000	-62.0816	-26.83822

Although it is not possible to make a robust decision for the patenting behavior of the network based on only three significant indicators, it could be inferred that the scientists who are involved in patenting activities do not commonly participate in article co-authorship relationships. In other words, scientists who tend to take advantage of their innovation commercially have fewer tendencies to get involved in the knowledge exchange and spillover process in the network. This conclusion is specifically inspired by the very high positive coefficient of the degree centralization suggesting that the high variation in the number of links among the scientists in the network favors the patenting of the network. Besides, since the coefficient for the path length is positive and relatively high, our conclusion is strengthened. However, further research is required to elucidate this issue.

4.2.3. Regression results for the *Total number of patent claims* model

The negative binomial regression model developed for evaluation of the effect of the network structure on the patent quality (measured by the total number of patent claims) shows no significant impact of the network indicators, since all the calculated *p* values are very much larger than the critical value of 0.01. This fact strongly implies that the network of article co-authorships does not give us enough evidence to assess the impact the network structure of scientists' relationship on the quality of registered patents in Canadian biotechnology sector. Table 8 illustrates the results of the regression analysis taken from STATA.

Table 8: Negative binomial regression results for the total number of patent claims model

PQ	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ConNet	-.1768547	.3748458	-0.47	0.637	-.9115391	.5578296
DC	-49.32449	81.82136	-0.60	0.547	-209.6914	111.0424
CloC	13.72718	5.903019	2.33	0.020	2.157477	25.29689
BC	9.282727	7.128996	1.30	0.193	-4.689848	23.2553
Ln_LC	-.9246334	.893048	-1.04	0.300	-2.674975	.8257086
PLC	5.314466	4.169893	1.27	0.202	-2.858373	13.48731
PL	-5.40012	4.759875	-1.13	0.257	-14.7293	3.929064
CC	-8.018557	5.569834	-1.44	0.150	-18.93523	2.898117
SW	.0002464	.0010402	0.24	0.813	-.0017924	.0022852
Ln_Scts	1.119249	2.024268	0.55	0.580	-2.848243	5.086741
_cons	6.57206	24.20172	0.27	0.786	-40.86244	54.00656

Chapter 5: Conclusions, Limitations and Future Study

5.1. Conclusions

This study explores the network of biotechnology scientists and inventors in Canada. Specifically, the purpose was to examine the relationship between the structural properties of the network (particularly small-world properties) and the research and innovative performance of scientists and inventors within the network.

This study contributed to the literature from different aspects. Although previous studies have explored the networks of firms or individuals in biotechnology, the Canadian biotechnology network was not their area of interest and this study is one of the first ones investigating this sector in Canada over a very long period of time.

Moreover, some of the studies have examined the research collaboration effect on the knowledge productivity of the networks; also the patent co-inventorship relationships have been studied to explore their influence on the patenting productivity of the inventors' network. But, to my knowledge, this is the first study that examined the impact of knowledge exchanges occurring during article co-authorship collaboration on the patenting productivity of inventors.

The first objective of this research was to quantify the small-worldliness of Canadian biotechnology network of scientists and to observe whether its structure resembles the small-world network structure or not. According to the results, the network under study showed significant small-world properties in any aspect i.e. the path length of the networks are very close to the ones of random networks; clustering coefficients of the networks are much larger than the corresponding values of clustering coefficients approximated for the random networks; and finally, the small-world ratios are great or even larger than the corresponding values of previously studied networks.

The other interesting finding of this study deals with the Milgram's (1967) claim regarding the notion of six degrees of separation. According to his study on average only six intermediates are needed to reach a person who is unknown to an individual. Our results strongly support Milgram's finding; according to this study, the separation degree between scientists converge to six in the networks built on the later time periods. Therefore, as our networks demonstrate more small-world characteristics, the number of intermediaries between individuals become closer to six.

Another research question of this study was related to the relationship between structural properties of co-authorship network of Canadian biotechnology scientists and their knowledge output. Our results proved that there is a significant association between the way the scientists are interconnected among themselves when collaborating on their research papers and the number of publications arising from these collaborations.

Next research questions investigated the relationship between the structure of the scientific network and the innovative performance in terms of both the quantity and the quality of the innovations generated by the inventors in this network. Based on the results we conclude that there is no great association between the pattern of knowledge exchange among the collaborating scientists in the network and the network's innovation productivity, whether assessed by the quantity or by the quality of the patents.

It should be mentioned that the model developed for analyzing the patent productivity supported a previously stated hypothesis that says the inventors who produce patents do not widely participate in co-authorship relationships. This could be a result of policies of firm owners and managers for security purposes and for maintaining their superiority advantages.

The last research question addressed in this study examined the possible effect of small-world structure of the network under study on the knowledge productivity and innovativeness of the whole network. Although our second and third models (patents and patent claims) demonstrated no effect on our dependent variables, the articles model outcomes were in accordance with the findings of some previous studies. The common hypothesis, which states that small-world properties enhance the knowledge creation, is partially supported in this study. The results show positive impact of small-world on the knowledge productivity of Canadian biotechnology sector.

The small-world effect was decomposed into the effect of shortest path length and the effect of cliquishness, where their impacts on the knowledge creation were studied separately. As expected, it was found that the shortest path length demonstrates positive effect on the scientific knowledge productivity. This is in consistence with the widely accepted findings that the short path length improves the information transfer in the network by enabling easier flow of knowledge among different network members. On the other hand, our results show that clustering coefficient has a negative impact on the research performance of the network. The role of clustering in networks has been analyzed by many researchers, whose conclusions are not consistent, because both positive and negative effects have been reported. Our work supports the finding that the high clustering of the network limits the knowledge creation due to the large amount of redundant information in that network, which is consistent with the outcomes of for example Fleming (2007), Gilsinget et al. (2008) and He (2009). Finally, we examined the magnitude of the effect for both small-world variables, and we conclude that the research productivity increase caused by shorter path length is greater than the decline in scientific performance resulting from the increased clustering coefficient.

5.2. Limitations of the study

We were exposed to some limitations in our analysis which are listed in this section.

First limitation is that the small-world factors (PL, CC) and the closeness centralization could be calculated on the largest connected component of the network, and not on the full network containing all the nodes. As a result, these measures may be biased. However, the greater is the proportion of the largest component, the lesser bias is involved. The largest components in our networks have outstanding increasing trend over the years, and in the last time intervals they cover almost 90% of the nodes in the network. Still, a certain level of noise in the models may persist. Although there have been some recommendations as to how to resolve this issue, the solutions are not usually applicable when the special purpose software for social network analysis is used. A solution proposed by Schilling and Phelps (2009) is further discussed in the following section. ,

Next, we were not able to assess the relationship between the structural properties of the articles co-authorship network and the quality of publications produced by the network members. The most commonly accepted measure for the quality of the research articles is the number of citations which each individual article receives from other citing research papers. However, since our data did not include this information, we could not examine this relationship.

The other limitation we faced is that many important factors that affect our dependent variables under study could not be incorporated in the models. For example, even though our methodology is able to detect and analyze the research collaborations leading to some tangible output (article, patent), the informal relationships that exist among scientists were completely neglected. These types of connections are never recorded and as a result could not be quantified, but there are certainly some knowledge exchanges occurring during such associations that could affect the network performance.

5.3. Future study

In this section of the thesis, some recommendations for future research are proposed. The first part of suggestions is related to the possible solutions of the current limitations explained in previous section. After that some other theoretical and methodological recommendations are proposed.

The most important limitation explained above is the inability to compute the small-world properties for the whole network, while considering only the largest connected component instead. An interesting solution for this problem was proposed by Schilling and Phelps (2009) i.e. considering the inverse of path length instead of the actual path length value. In the previous methods, the distance of two nodes from separate components would be infinite, but in their method the ratio will be zero. They called it the distance weighted reach and described it for a firm in a network as: "A firm's distance-weighted reach is the sum of the reciprocal distances to every firm that is reachable from a given firm, i.e. $\sum_j \frac{1}{d_{ij}}$, where *dij* is defined as the minimum distance (geodesic), d, from a focal firm i to partner j , where i ≠ j. A network's average distance-weighted reach is this measure averaged across all firms in the network, $(\sum_n \sum_j \frac{1}{d_{ij}})/n$, where n is the number of firms in the network."

Using this method, the small-world ratio could be calculated over the whole network. The same approach could be applied to the closeness centralization. However, it should be noticed that this method is not available in the current social network analysis software and needs specialized programming.

Another interesting factor to be considered in the analysis is the strength of the association between scientists. So far, multiple connections between individuals were considered as a single link in our networks. It is thus proposed here to consider the number of articles coauthored by the scientists during each time interval as a measure of the strength of their relationship. Hence, further research is needed to investigate the change in the model outcome after the inclusion of the strength of the relationships. Some light will also be shed on the change in the impact of other structural indicators on the network performance.

Moreover, we also propose to include international scientific relationships in the analysis. Current study takes into account only Canadian scientists, but we also detected numerous connections between them and other biotechnology scientists outside Canada. The connections to these international researchers create important channels for knowledge originating in highly research intensive areas (especially in the USA), through which the knowledge is transmitted and finally received here, in Canada. This knowledge is very valuable for Canadian researchers, since it involves new and fresh information and ideas that can significantly enhance the creativity of the Canadian scholars and inventors. Furthermore, since we assume that the number of connections between local scientists and the ones from more distant organizations is lower than the number of the links within Canada, it is expected that the long range links will act as bridges and the small-world properties will be intensified.

In addition, it is suggested to include the research collaboration relationships at the firm level as well. In our analysis, only the connections among individuals have been assessed, while most of the literature evaluated only the firm level networks. By taking into account the affiliations of the scientists, the inter-firm collaboration networks could be constructed and considered in the analysis together with our existing networks of individual scientists. Analyzing both networks at the same time may bring further insights into the problem and can achieve a great contribution in this field. Finally, we also propose to analyze the effect of network structure on the performance of the network while employing various time lags. In the current study, the dependent variables are taken from the subsequent years of the time intervals, as it was assumed that this is when the final outcome of the collaboration is realized. It can be expanded to more models by considering the productivity of the network in two or three-year lag perspectives.

Finally, since our findings demonstrates no significant impact of the article co-authorships network on the patenting performance and the quality of patents, it is recommended that future researchers focus more on the patent co-authorship networks instead of article collaboration relationships to analyze and improve its affect on the innovativeness of the network.

References

Abrahamson, E., & Rosenkopf, L. (1997). Social network effects on the extent of innovation diffusion: A computer simulation. *Organization science*, 289–309.

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative science quarterly*, 425–455.

Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, *74*(1), 47.

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.

Baum, J. A. C., Shipilov, A. V., & Rowley, T. J. (2003). Where do small worlds come from? *Industrial and Corporate change*, *12*(4), 697.

Beaudry, C., & Schiffauerova, A. (2011). Impacts of collaboration and network indicators on patent quality: The case of Canadian nanotechnology innovation. *European Management Journal*.

Bland, C. J., Ruffin, M. T., & others. (1992). Characteristics of a productive research environment: literature review. *Academic Medicine*, *67*(6), 385.

Brantle, T. F. (2011). *Complexity, innovation and economic growth: The competitive network of innovation and organizational size and growth in innovation*. STEVENS INSTITUTE OF TECHNOLOGY.

Brantle, T. F., & Fallah, M. H. (2007). Complex innovation networks, patent citations and power laws. *Management of Engineering and Technology, Portland International Center for* (pp. 540–549).

Carayol, N., & Roux, P. (2005). Self-Organizing Innovation Networks: when do small worlds emerge? *European Journal of Economic and Social Systems*, *18*(2), 307.

Centra, J. A. (1983). Research productivity and teaching effectiveness. *Research in Higher Education*, *18*(4), 379–389.

Choi, H., Kim, S. H., & Lee, J. (2010). Role of network structure and network effects in diffusion of innovations. *Industrial Marketing Management*, *39*(1), 170–177.

Clements. M. M. (2008). *Patenting at universities in the United States: A network analysis of the complexities of domestic and international university patenting activities* (Ph.D., School of Education). Indiana University, United States -- Indiana.

Coleman, J. S. (1981). *Longitudinal data analysis*. 102-081-920.

Cowan, R., & Foray, D. (1997). The economics of codification and the diffusion of knowledge. *Industrial and corporate change*, *6*(3), 595–622.

Cowan, R., & Jonard, N. (2004). Network structure and the diffusion of knowledge. *Journal of economic Dynamics and Control*, *28*(8), 1557–1575.

Davis, G. F., Yoo, M., & Baker, W. E. (2003). The small world of the American corporate elite, 1982-2001. *Strategic organization*, *1*(3), 301–326.

De Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory social network analysis with Pajek* (Vol. 34). Cambridge Univ Pr.

Demirkan. (2007). *Essays on the evolution of networks in the U.S. biotechnology industry* (Ph.D.). The University of Texas at Dallas, United States -- Texas.

Dolfsma, W., & Leydesdorff, L. (2008). Innovation systems as patent networks. *Conference of European Association for Evolutionary and Political Economics, Rome* (pp. 5–7).

Fleming, L., King, C., & Juda, A. I. (2007). Small Worlds and Regional Innovation. *Organization Science*, *18*(6), 938 -954. doi:10.1287/orsc.1070.0289

Gilsing, V., Nooteboom, B., Vanhaverbeke, W., Duysters, G., & Van Den Oord, A. (2008). Network embeddedness and the exploration of novel technologies: technological distance, betweenness centrality and density. *Research Policy*, *37*(10), 1717–1731.

Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological theory*, *1*(1), 201–233.

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 1360–1380.

Hargadon, A. (2003). *How breakthroughs happen: the surprising truth about how companies innovate*. Harvard Business Press.

Hausman, J. A., Hall, B. H., & Griliches, Z. (1984). *Econometric models for count data with an application to the patents-R&D relationship*. National Bureau of Economic Research.

He, J., & Fallah, M. H. (2006). Mobility of Innovators and Prosperity of Geographical Technology Clusters: A longitudinal examination of innovator networks. *Proceedings of International Conference on Complex System, Boston, MA, June* (pp. 24–30).

He, J., & Hosein Fallah, M. (2009). Is inventor network structure a predictor of cluster evolution? *Technological forecasting and social change*, *76*(1), 91–106.

Jaffe, A. B., Trajtenberg, M., & Fogarty, M. S. (2000). Knowledge spillovers and patent citations: Evidence from a survey of inventors. *American Economic Review*, *90*(2), 215–218.

Katz, E. (1961). The social itinerary of technical change: two studies on the diffusion of innovation. *Human Organization*, *20*(2), 70–82.

Kogut, B., & Walker, G. (2001). The small world of Germany and the durability of national networks. *American Sociological Review*, 317–335.

Kuzhabekova. A. (2011). *Impact of co-authorship strategies on research productivity: A social-network analysis of publications in Russian cardiology* (Ph.D., Educational Policy and Administration). University of Minnesota, United States -- Minnesota.

Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters*, *87*(19), 198701.

Li, Y., & Chen, Z. (2007). Diffusion of Innovations in a Small World Network. *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on* (pp. 5617–5620).

Liebeskind, J. P. (1996). Knowledge, strategy, and the theory of the firm. *Strategic management journal*, *17*(WINTER), 93–107.

Liu, B. S. C., Madhavan, R., & Sudharshan, D. (2005). DiffuNET: The impact of network structure on diffusion of innovation. *European Journal of Innovation Management*, *8*(2), 240–262.

McFadyen M. Ann, Semadeni M., Cannella A. A. (2009). Value of Strong Ties to Disconnected Others: Examining Knowledge Creation in Biomedicine. *Organization science Vol. 20, No. 3, May–June 2009, pp. 552–564*

Melin, G., & Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics*, *36*(3), 363–377.

Midgley, D. F., Morrison, P. D., & Roberts, J. H. (1992). The effect of network structure in industrial diffusion processes. *Research Policy*, *21*(6), 533–552.

Milgram, S. (1967). The small world problem. *Psychology today*, *2*(1), 60–67.

Monasson, R. (1999). Diffusion, localization and dispersion relations on “small-world” lattices. *The European Physical Journal B-Condensed Matter and Complex Systems*, *12*(4), 555–567.

Newman, M. E. J. (2000). Models of the small world. *Journal of Statistical Physics*, *101*(3), 819–841.

Newman, M. E. J. (2001a). Clustering and preferential attachment in growing networks. *Physical Review E*, *64*(2), 025102.

Newman, M. E. J. (2001b). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, *98*(2), 404.

Newman, M. E. J. (2001c). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, Phys. Rev. E, Stat. Nonlinear Soft Matter Phys. (USA), *64*(1), 016131-1. doi:10.1103/PhysRevE.64.016132

Newman, M. E. J. (2001d). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, Phys. Rev. E, Stat. Nonlinear Soft Matter Phys. (USA), *64*(1), 016132-1. doi:10.1103/PhysRevE.64.016131

Newman, M. E. J. (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.

Newman, M. E. J., & Watts, D. J. (1999). Scaling and percolation in the small-world network model. *Physical Review E*, *60*(6), 7332.

Newman, M. E. J., Watts, D. J., & Strogatz, S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(Suppl 1), 2566.

Oliver, A. L., & Liebeskind, J. P. (1997). Three levels of networking for sourcing intellectual capital in biotechnology: implications for studying interorganizational networks. *International Studies of Management & Organization*, *27*(4), 76–103.

Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative science quarterly*, 116–145.

Powell, W. W., White, D. R., Koput, K. W., & Owen-Smith, J. (2005). Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences1. *American journal of sociology*, *110*(4), 1132–1205.

Reagans, R., & McEvily, B. (2003). Network structure and knowledge transfer: The effects of cohesion and range. *Administrative Science Quarterly*, 240–267.

Rumsey-Wairepo. A. (2006). *The association between co-authorship network structures and successful academic publishing among higher education scholars* (Ph.D.). Brigham Young University, United States -- Utah.

Sandström, A., Pettersson, I., & Nilsson, A. (2000). Knowledge production and knowledge flows in the Swedish biotechnology innovation system. *Scientometrics*, *48*(2), 179–201.

Schank, T., & Wagner, D. (2004). *Approximating clustering-coefficient and transitivity*. Universität Karlsruhe, Fakultät für Informatik.

Schiffauerova, A., & Beaudry, C. (2008a). Collaboration and network indicators in Canadian nanotechnology. invited presentation**,** *National Bureau of Economic Research (NBER) Conference on Emerging Industries: Nanotechnology and NanoIndicators,* Cambridge, Massachusetts, May 1-2.

Schiffauerova, A., & Beaudry, C. (2008b). Innovation Networks and Collaboration in Canadian Nanotechnology Clusters. *The 42nd Pacific Northwest Regional Economics Conference*, The Spirit of Innovation Forum, Tacoma, Washington, May 14-16.

Schiffauerova, A., Beaudry, C. (2009). Canadian nanotechnology innovation networks: intra-cluster, inter-cluster and foreign collaboration. *Journal of Innovation Economics*, (2), 119–146.

Schilling, M. A., & Phelps, C. C. (2004). Interfirm collaboration networks: The impact of small world connectivity on firm innovation. *Unpublished manuscript*.

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.

Taylor, M. S., Locke, E. A., Lee, C., & Gist, M. E. (1984). Type A behavior and faculty research productivity: What are the mechanisms? *Organizational Behavior and Human Performance*, *34*(3), 402–418.

Tong, X., & Frame, J. D. (1994). Measuring national technological performance with patent claims data. *Research Policy*, *23*(2), 133–141.

Toutkoushian, R. K., Porter, S. R., Danielson, C., & Hollis, P. R. (2003). Using publications counts to measure an institution's research productivity. *Research in Higher Education*, *44*(2), 121–148.

Travers, J., & Milgram, S. (1969). An experimental study of the small world problem. *Sociometry*, 425–443.

Uzzi, B. (2008). A social network's changing statistical properties and the quality of human innovation. *Journal of Physics A: Mathematical and Theoretical*, *41*, 224023.

Uzzi, B., & Spiro, J. (2005). Collaboration and Creativity: The Small World Problem1. *American Journal of Sociology*, *111*(2), 447–504.

Valente, T. W. (1996). Network models of the diffusion of innovations. *Computational & Mathematical Organization Theory*, *2*(2), 163–164.

Wang, N., Wang, X., Sun, Q., & Zhao, L. (2008). Innovations Diffusion on Heterogeneous Newman-Watts Small-World Network: A Computer Simulation. *Information Management, Innovation*

Management and Industrial Engineering, 2008. ICIII'08. International Conference on (Vol. 1, pp. 296–299).

Wasserman, S., & Faust, K. (1994). *Social network analysis: methods and applications*. Cambridge University Press.

Watts, D. J. (1999). Networks, dynamics, and the small-world phenomenon 1. *American Journal of Sociology*, *105*(2), 493–527.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world"networks. *Nature*, *393*(6684), 440–442.

Zucker, L. G., Darby, M. R., Brewer, M. B., & Peng, Y. (1995). *Collaboration structure and information dilemmas in biotechnology: Organizational boundaries as trust production*. National Bureau of Economic Research.

Appendix A: Examples of two-mode and one-mode networks of Canadian biotechnology scientists

In this figure the red nodes represent the scientists and green nodes denote the articles. Lines connect articles to their authors.

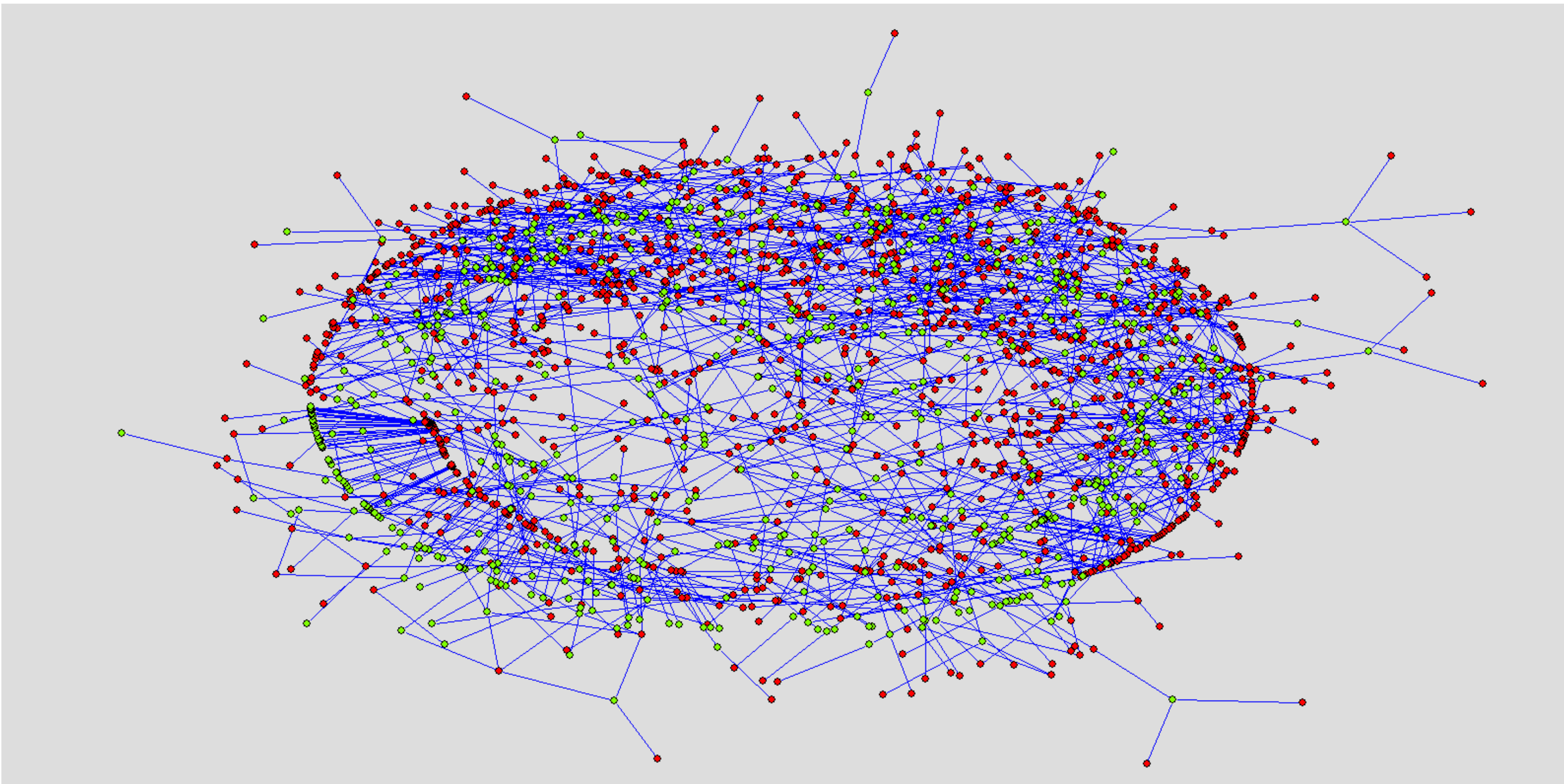

Figure 16: Two-mode affiliation network of Canadian biotechnology scientists from 1969 to 1973

In this figure scientists are connected to their collaborators. Therefore each link represents one (or more) article co-authorship.

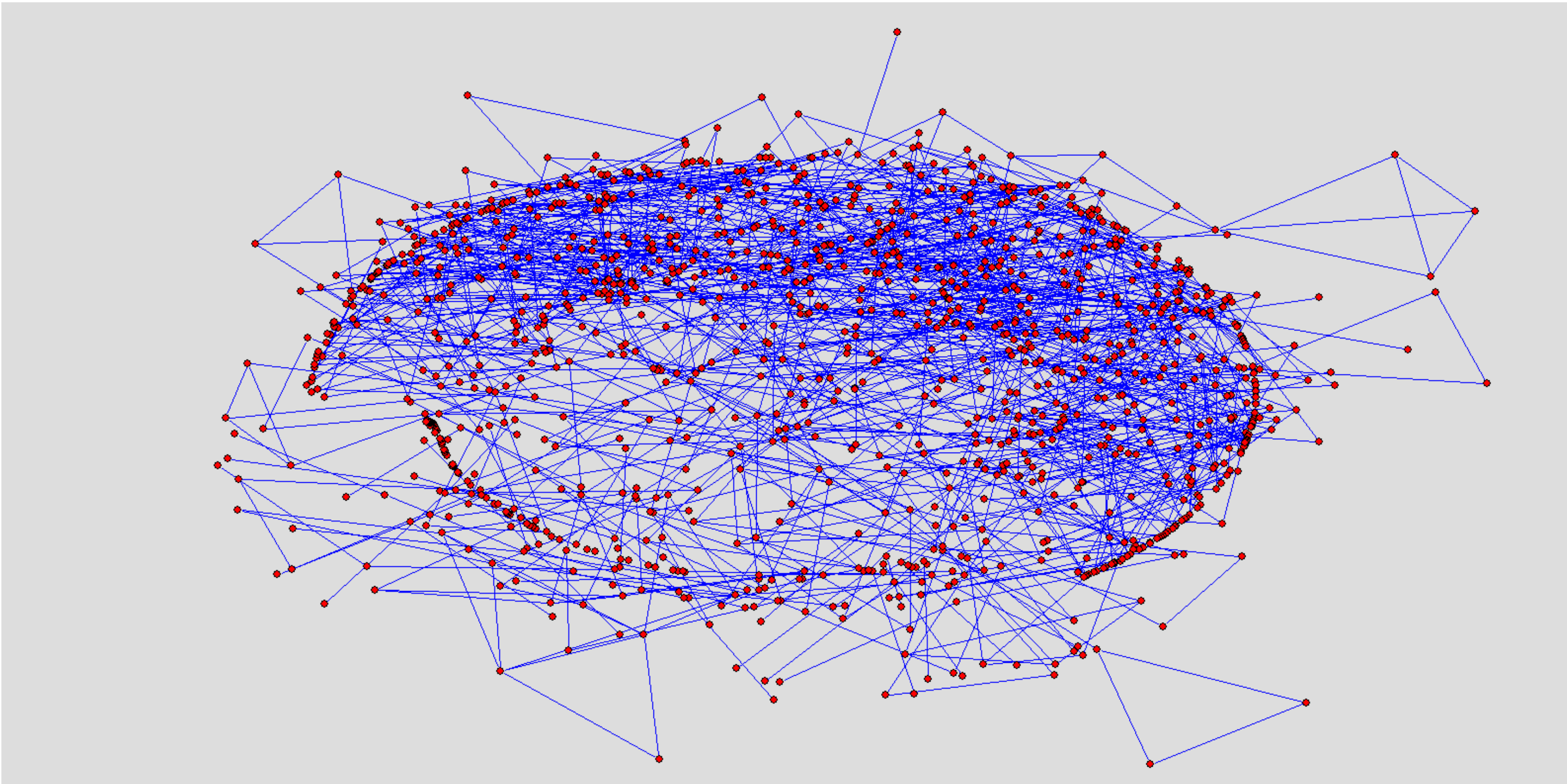

Figure 17: One-mode network of Canadian biotechnology scientists from 1969 to 1973

Appendix B: Results of the likelihood-ratio test for the over dispersion coefficient (Alpha)

Table 9: Likelihood-ratio test for over dispersion coefficient of articles model

ARTC	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ConNet	.0402313	.0383244	1.05	0.294	-.0348832	.1153458
DC	-25.57747	8.641421	-2.96	0.003	-42.51434	-8.640595
CloC	1.549786	.6898726	2.25	0.025	.1976603	2.901911
BC	-.1110232	.7544474	-0.15	0.883	-1.589713	1.367667
Ln_LC	.05364	.093327	0.57	0.565	-.1292776	.2365575
PLC	-.4026344	.4257402	-0.95	0.344	-1.23707	.4318011
PL	-1.333065	.5130039	-2.60	0.009	-2.338534	-.3275956
CC	-2.142378	.6282228	-3.41	0.001	-3.373672	-.9110842
SW	.0003269	.0001018	3.21	0.001	.0001273	.0005264
Ln_Scts	.4161176	.2059963	2.02	0.043	.0123723	.8198629
_cons	6.083844	2.445049	2.49	0.013	1.291636	10.87605
/lnalpha	-6.944179	.2763095			-7.485736	-6.402623
alpha	.0009642	.0002664			.000561	.0016572

Likelihood-ratio test of alpha=0: chibar2(01) = 304.22 Prob>=chibar2 = 0.000

Table 10: Likelihood-ratio test for over dispersion coefficient of patents model

P	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ConNet	-.4542267	.3229803	-1.41	0.160	-1.087256	.1788031
DC	56.10588	77.70414	0.72	0.470	-96.19144	208.4032
CloC	5.115553	5.37177	0.95	0.341	-5.412923	15.64403
BC	4.089766	6.543241	0.63	0.532	-8.734752	16.91428
Ln_LC	-.1332366	.8067274	-0.17	0.869	-1.714393	1.44792
PLC	1.472398	3.727974	0.39	0.693	-5.834297	8.779094
PL	1.790201	4.513377	0.40	0.692	-7.055855	10.63626
CC	1.439614	5.604236	0.26	0.797	-9.544487	12.42371
SW	-.0001459	.0008243	-0.18	0.859	-.0017616	.0014698
Ln_Scts	2.819212	1.77759	1.59	0.113	-.6647999	6.303224
_cons	-22.0999	20.97403	-1.05	0.292	-63.20824	19.00844
/lnalpha	-3.04344	.3302822			-3.690781	-2.396099
alpha	.0476706	.0157448			.0249525	.0910726

Likelihood-ratio test of alpha=0: chibar2(01) = 151.16 Prob>=chibar2 = 0.000

Table 11: Likelihood-ratio test for over dispersion coefficient of patent quality model

PQ	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ConNet	-.1768547	.3748458	-0.47	0.637	-.9115391	.5578296
DC	-49.32449	81.82136	-0.60	0.547	-209.6914	111.0424
CloC	13.72718	5.903019	2.33	0.020	2.157477	25.29689
BC	9.282727	7.128996	1.30	0.193	-4.689848	23.2553
Ln_LC	-.9246334	.893048	-1.04	0.300	-2.674975	.8257086
PLC	5.314466	4.169893	1.27	0.202	-2.858373	13.48731
PL	-5.40012	4.759875	-1.13	0.257	-14.7293	3.929064
CC	-8.018557	5.569834	-1.44	0.150	-18.93523	2.898117
SW	.0002464	.0010402	0.24	0.813	-.0017924	.0022852
Ln_Scts	1.119249	2.024268	0.55	0.580	-2.848243	5.086741
_cons	6.57206	24.20172	0.27	0.786	-40.86244	54.00656
/lnalpha	-2.262849	.2545469			-2.761752	-1.763947
alpha	.1040536	.0264865			.063181	.1713672

Likelihood-ratio test of alpha=0: chibar2(01) = 2756.05 Prob>=chibar2 = 0.000