The Role of Individuals in Innovation Networks: A Simulation Approach in Canadian Biotechnology Network

Dorsa Tajaddod Alizadeh

A Thesis in the Department of Mechanical and Industrial Engineering

Presented in Partial Fulfillment of the Requirements for the Degree of Master of Applied Science

(Industrial Engineering) at Concordia University

Montreal, Quebec, Canada

December 2001

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Dorsa Tajaddod Alizadeh

Entitled: The Role of Individuals in Innovation Networks: A Simulation Approach in Canadian Biotechnology Network

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

Master of Applied Science(Industrial Engineering)

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

Signed by the final examining committee:

Dr. Onur Kuzgunkaya ______ Chair

Dr. Simon Li ______ Examiner

Dr. Majlinda Zhegu ______ Examiner

Dr. Andrea Schiffauerova ______ Supervisor

Approved by ______
Chair of Department or Graduate Program Director

______
Dean of Faculty

Date May 1, 2012

Abstract

The Role of Individuals in Innovation Networks: A Simulation Approach in Canadian Biotechnology Network

Dorsa Tajaddod Alizadeh

Today, innovation is the key to survival in biotechnology markets. Innovation occurs in collision of different sources of knowledge in innovation networks. The innovation networks have already attracted the scholarly attention, but the research remains mostly at the firm and cluster levels, while much less research effort has been devoted to the study of the roles of individuals and their relationships. Moreover, the innovation networks are usually investigated as static, while the study on their dynamics has generated much less interest among the researchers. The contribution of this thesis is twofold. First, a simulation model of Canadian biotechnology innovation networks is developed using real data on the publications and patents extracted from the USPTO and Scopus databases. This scientists-level simulation model has been created as very flexible, and as such it can be used for further research examining the Canadian biotechnology network under various conditions. Second, the thesis has investigated the roles of individual scientists and their relationships in terms of the network innovative productivity and its knowledge transmission capability. While the repetitiveness of the collaborative relationships among scientists has shown quite negative effects, the presence of the Gatekeepers has proved to be very positive for the overall efficiency of the innovation network. The impact of star scientists on the innovative activities has been found positive, but some negative effects on the flow of knowledge in the network have

been detected. Finally, we make recommendations for future research using the developed tool for the study of the dynamics of the innovation networks.

Acknowledgements

It is a pleasure for me to offer my sincerest gratefulness to those who provided me with help and support to make the completion of this thesis possible.

First and foremost, I owe my deepest gratitude to my parents, and my dearest brother, who have always granted me the truest love, purest energy, and priceless support, without which I could not accomplish any of my goals in life up to the present.

Besides, I would like to show my thankfulness to my supervisor, Professor Andrea Schiffauerova, for her precious directions, guidance, and assistance. It was my best of luck to have her innovative ideas and helpful inspirations throughout the entire procedure of my research.

I do express my special thanks to my significant other, Hamidreza, without whom I would have got lost far before now. I also would like to show my appreciation to my dear friends Afshin and Maryam, who supported me with their priceless technical abilities.

Finally, I do appreciate Concordia University for this unique opportunity.

Table of Contents

List of Tables

List of Figures

1. Introduction

As customer demand for new products grows all over the world, the traditional methods of production and services are becoming less reliable. Today, firms are looking for many new and specialized products and procedures to survive in the current highly competitive world of commerce. The principal key to succeed in fulfilling the radically changing demand of customers is nowadays believed to be innovation.

Today, innovativeness goes beyond the borders of single industries, and therefore firms should try to enlarge their external knowledge sources of relevant technologies. The increasing pace of knowledge generation, as well as the decreasing life cycle of innovative products/procedures, has forced the firms to look for the situations in which they can have a quick access to new knowledge.

Knowledge is primarily produced and certified in the laboratories and research branches of any establishment, such as universities, research and development sections of the firms, governmental research programs, etc. Once some piece of new knowledge is achieved, the beholders of the knowledge will try to utilize it to attain their objectives. However, creation of new knowledge is both expensive and time consuming, and requires specialized people and vast access to the information sources. Therefore, it is more reasonable to have continuous interactions with the knowledge possessors in order to gain the advantages of the new knowledge they generate. (Robertson and Langlois 1995)

In this regard, involvement in the innovation networks[1] is considered as a good solution for firms to get updated knowledge and information they need. As firms and universities start to collaborate through innovation networks, more and more flows of information are created that connect innovators in various firms in the network.

The mechanisms of knowledge creation and also learning processes are playing the most important role in the innovation networks. The operational behaviour of networks has been studied in the last decades, in order to find applicable theories, rules and policies for firms and governments to accelerate the innovative efforts. However most of the research in this regard is mainly at the observational and description level (Zaheer and Bell 2005; Rogers 1995; Becheikh, Landry, and Amara 2006; Gopalakrishnan and Damanpour 1997; Porter 1998; Zucker, Darby, and Armstrong 1998, and many others) and only a few have tried to examine the theories and inspect the effectiveness rate of innovation networks (Gilbert 2004; Albino, Carbonara, and Giannoccaro 2006; Pyka, Gilbert, and Ahrweiler 2009).

There are yet a number of questions that must be answered regarding the dynamic behaviour of innovation networks, and the role of individual scientists in transmission of knowledge, so as to facilitate the procedure of managing and controlling the networks, and also to allow firms make new regulations and decisions regarding innovation networks and flow of knowledge.

The individual scientists in the networks have different levels of importance according to their position and capabilities regarding knowledge production and transmission. The

[1] In this thesis, innovation network refers to any kind of social network in which new knowledge is produced by the nodes of the network and transferred between them.

inventors, who connect many directly unconnected clusters[2] together, are called Gatekeepers. As the innovation networks grow, the role of the Gatekeepers becomes more essential for the solidarity of the network and its survival, since they prevent the networks from division.

Star scientists are also another class of scientists, the role of whom affect the structure, productivity and improvement of the innovation networks significantly. Generally, stars are the scientists with noticeably higher number of patents and publications. In other words, star scientists are those who utilize the knowledge flow in the networks more efficiently, and produce more new knowledge than others.

Besides the specific roles of individuals in the innovation networks, the extent of loyalty to previous partners, i.e. the willingness of scientists to work with their earlier collaborators, rather than searching for new partners may affect the network's structure and productivity. Therefore, it would be interesting to study the impacts of the loyalty of scientists on the performance of the innovation networks, in order to come up with better understanding of the effects of mutual relationships on innovation.

The innovation networks are really complicated, and not all of the internal interactions are easy to recognize and study at the time. Therefore, simulation models are the best solutions to imitate the behaviour of the networks and to copy their characteristics, so the researchers can examine the validity of their hypotheses and theories. With the assistance of computers, the simulation based studies about innovation networks are increasing recently. The concept of agent-based modeling has advanced the learning capabilities of

[2] In this thesis, the word cluster refers to any partitioning of nodes in the network, which can be based on their various characteristics such as geographical situation, organization type, field of work, etc.

researchers for investigating various aspects of networks, and examining theories, rules and regulations that could improve the efficiency of the networks.

The research objective of the present thesis is to employ simulation modeling methodology in the domain of innovation networks in order to develop a flexible simulation model, and reveal the related features of the Canadian biotechnology network in a scientist-level of study, and analyse the model for various scenarios of the presence and absence of the Gatekeepers, and star scientists, and different levels of loyalty in mutual relationships of individual scientists, and find appropriate solutions and answers to the research questions of the thesis.

In the next part of the thesis, the relative literature on the innovation networks and their characteristics is reviewed.

2. Literature review

In this part of the thesis the relevant research work from the literature is reviewed. The available literature can be divided into three main categories: the first category introduces the literature on the existence of industrial clusters and innovation network, and the related definitions, characteristics and factors. In the second category the simulation based surveys and their approach and results are reviewed. Finally, the last part reviews the literature about the role of the Gatekeepers, star scientists and individual scientists in the innovation networks.

2.1 Innovation

Innovation can be defined as any alteration of thought, process, product or manner that leads to a better condition, i.e. increased satisfaction, productivity, wealth, etc. Because of the broad meaning of innovation and its applicability to almost any field of science, namely engineering, economics, psychology, sociology, marketing, etc. it has been addressed many times in numerous research works. Although all of the researchers are thinking about "newness" when issuing innovation, the viewpoints are diversified in various fields. This survey's objective is to cover innovation mostly in the perspective of economics and technology management.

"Innovation plays a role in nurturing the economy, in enhancing and sustaining the high performance of firms, in building industrial competitiveness, in improving the standard of living, and in creating a better quality of life", as Gopalakrishnan and Damanpour (1997), and others state. Therefore, understanding its nature and emergence procedure in manufacturing firms and industries can assist managerial activities to make operative decisions that will foster innovativeness, which is greatly emphasized in recent research works (Lenz-Cesar and Heshmati 2010; Harrisson and Laberge 2002; Pittaway et al. 2004).

Innovation can help companies in two distinct manners: introducing new products with new features to satisfy demands more specifically; or developing new procedures of manufacturing products to compete with rivals on the rate of responsiveness, efficiency and cost reduction.

In the book "The theory of economic development", 1934, Joseph Schumpeter argues that if a firm wants to maintain its competitive advantage on the market, attract customer satisfaction, or enter a new market, innovation should be set as its priority. Firms should attempt to be innovative in all of their fundamental activities, from transportation to employment policies, and learning innovation essentials and features could help them in this regard.

In the literature, innovation has always been considered within two distinct categories: 1- Discrete outcome of innovation in form of either a new product or a new procedure, arising as a new idea, methodology or device (Kimberly and Evanisko 1981; Damanpour and Evan 1984); 2- The innovation development procedure (VAN DE VEN and Rogers 1988; Ettlie 1980). The first category is mostly related to the marketing aspect of the innovation, and the adoption of innovation after it is introduced. However, in this survey the second aspect of innovation is taken into account, which deals with the procedure before and during the time when innovation happens.

Either by improving existing products and procedures, or by introducing something totally new, innovation is nowadays proved to increase the competitive advantage of companies in various industries. With the rapidly changing demands of clients for innovative products, it is vital for firms to have a vast innovation capability to survive in the current, rapidly-changing business world (Udell, Bottin, and Glass 1993; Lenz-Cesar and Heshmati 2010).

Although the importance of innovation is already known, and its advantages are already vastly discussed, there have been few researches done on the elements that determine innovation development in the firms. There exist only a few empirical studies that specify the factors with key role in the development of innovation in various industries (Frenken 2000; Gilbert, Ahrweiler, and Pyka 2007; Shih and Chang 2009).

At the end of the 19^{th} century, the innovation networks and their principles attracted the scholarly attention increasingly. Souitaris (1999, 2001, and 2002) has tested the effect of external communications on the innovative behaviour of firms, and also the effect of firm and the region on this behaviour, through empirical studies. Many others have also done related studies that specify the affecting factors on innovation in organizations (Acs and Audretsch 1987), effect of firm size; (Galende and Fuente 2003), internal resources and factors effect; (Nohria and Gulati 1996), slack impact on innovation; (Damanpour 1987), organizational factors; (Laursen and Foss 2003), human resource and management effects; etc.). In all of these studies one or more variables have been taken into consideration and various degrees of association with innovation rate have been discovered. However, Coombs et al. (1996) believe that this kind of literature does not help in comprehending the innovation phenomenon as a whole.

In almost all of the studies, the key factor in the procedure of developing innovation is said to be knowledge. In other words, innovation is achievable by experimenting available knowledge in unexplored ways, by practicing new knowledge, or by using a combination of available and new knowledge. There are two distinct methods for achieving new knowledge: either by using inter-firm R&D departments, or by exploiting the overflow of information from external sources like other firms, universities, research departments, etc.

However, having access to new knowledge alone is not enough to gain creativity in products and to satisfy the customer requirements. There are other factors that affect the development of product innovations, related to the way that knowledge transfers among various sources. The term "knowledge transfer" and its characteristics have been known as a fundamental factor for improving competitive advantage of companies since long time ago. In this regard, the role of "innovation networks", and "industrial clusters", also known as "industrial districts (ID)", have been discussed in various studies, e.g. Marshall (1890), Krugman (1991), Porter (1998), etc.

Conclusion: The literature targets the innovation as a fundamental necessity for improvement in the markets. The innovation is said to occur in the collision of different sources of new or available knowledge. Although innovation has been already vastly discussed by various authors, its principles and characteristics yet must be studied more by analyzing the elements that would affect the transmission of knowledge inside the innovation networks. In the next section, the related literature on the definitions of clustering and innovation networks is covered, and the related features of innovations affecting the circulation of knowledge are reviewed.

2.2 Clusters: pros and cons for innovation

Any group of competing/co-operating linked firms, companies and other related institutions that are close geographically can be thought of as a cluster (Porter 1998). Although the concept of "cluster" was first introduced in the last decade, there are yet many different definitions for that found in various references (Martin and Sunley 2003).

In this thesis, the word cluster applies to a number of firms, organizations and institutions that are connected to each other within a geographical proximity, sharing benefits of the same market, and cooperating in knowledge exchange process, in order to produce innovative ideas that could suit the purposes of the whole cluster. The individual scientists from different clusters and with various research fields are then nodes of the network, which can produce, transmit, or absorb knowledge as their interactions with each other trough co-publications.

Marshall (1890) seems to be the first one bringing up the theories on clustering, and later Krugman (1991) shed some light on the procedures and characteristics of the clusters. Krugman claims that the higher availability rate of labour and intermediate inputs, as well as better chances of exposure to knowledge spillovers[3] attract new companies to enter the clusters, and result in the evolvement of the clusters.

The role of the industrial clusters in the innovation growth is highly dependent on the nature of the industry in which they are studied. In the knowledge-intensive industries, such as biotechnology or nanotechnology, the more connections a firm establishes in the

---

[3] Zucker *et al.* (1998) define knowledge spillovers as "Positive externalities of scientific discoveries on the productivity of firms which neither made the discovery themselves, nor licensed its use from the holder of intellectual property rights".

network, the more innovative it will become due to the high rate of knowledge generation, transition and spillovers inside the clusters. However, in the industries with less sensibility to knowledge, such as paper manufacturing industry, the innovativeness level of the firms does not alter significantly as their connections with other similar firms increase. (Baptista and Swann 1998)

The evolvement of the industrial clusters is mostly indebted to the centralization of the market, mediators, sources, and knowledge that attract the newcomers to join the cluster and benefit from its advantages. There are many benefits for the clusters mentioned in the literature, such as pooling the specialized people together, presence of specialized suppliers that could provide dedicated tools and services for the industry, lower costs of research activities due to the centralization of the knowledge within the clusters, lower transportation costs and wider focus on the whole market, and most importantly the information flow between the agents, also called spillover, which facilitates the innovative actions within the clusters. (Marshall 1961; Porter 1998; Prevezer 1997; Zucker, Darby, and Armstrong 1998 and many others)

Besides all the benefits of clustering, there are also some inevitable drawbacks. Growing competition between the firms in the cluster, declining pricing powers on the market, shortage and high prices of resources, loss of trade superiority, decreased profits, and shorter lifecycles of products and services are some of the disadvantages of the clusters discussed in the literature. (Meyer-Stamer 2002; Baptista and Swann 1998)

A successful cluster will result in faster growth of its firms, and higher rates of innovation, which in turn will attract more new firms to enter the cluster and lead to its own growth (Porter 1998). This has been also confirmed by other authors as well (Swann and Prevezer 1996; Baptista and Swann 1998). Moreover, they conclude that firms tend to grow faster in the clusters of their own industries. Besides, the stronger a cluster is, the faster firms grow in it (Baptista and Swann 1999).

The localized knowledge spillovers existing in the clusters have been suggested to help the firms located within the cluster to be more innovative than their rivals elsewhere (Dahl and Pedersen 2004). The knowledge generated inside the boundaries of a cluster can spread inside it more rapidly, and therefore firms inside the clusters are more capable to benefit from it. In this regard, both the position and proximity to the firms within the cluster plays an important role.

Conclusion: The benefits of the clustering have proven to overcome its drawbacks empirically (Cowan and Jonard 2004), and nowadays the size of the clusters is growing in different fields of industry all over the world due to the transmission of knowledge (Beaucage and Beaudry 2006). Innovativeness, as one of the main results of clustering, leads to the appearance of more and more individual innovators in the clusters (Rampersad, Quester, and Troshani 2010). These individuals are then playing the key role in producing and transferring the knowledge within the clusters. The connection of the individual innovators inside and outside the clusters will shape a network of innovators, called innovation network.

2.3 Innovation networks

Powell et al. (1996) indicate that learning processes can no longer remain inside the territories of the individual firms, and more attention should be paid to the inter-organizational collaborations, inter-firm activities, and their features and characteristics. In today's knowledge intensive economy, a firm needs to exploit knowledge resources more creatively and intelligently than its competitors, and establish a continuous learning procedure as well as cooperative research capabilities, in order to improve gradually (Cowan and Jonard 2003). A firm, even with highest levels of research activities, cannot survive on the market without cooperating with other firms and using available knowledge in related industries.

Dahl and Pedersen (2004) declare that high innovation rates and fast collection of knowledge, derived from the disclosure of information among competing agents, are created by groups of socially connected individuals. The socially connected groups of individuals are one of the direct key characteristics of innovation networks. Therefore, more information about the behaviour of networks and their structure, which lead to the transition of knowledge, is required to improve the performance of the whole industry (Cowan and Jonard 2003).

Zaheer and Bell (2005) argue that involvement in the innovation networks not only exposes firms to the outside sources of information and improves their innovativeness levels, but also enhances the productivity of their internal capabilities and boost their inner functionality efficiently.

There are many studies that discuss the importance of industrial collaborations, which emphasize the dependence of innovation improvement on the presence of innovation networks, and knowledge flows between the individuals from industrial groups and other sources of knowledge and information (Henderson and Cockburn 1996; Carneiro 2001; Midgley, Morrison, and Roberts 1992; Aharonson, Baum, and Feldman 2004). However, most of these works have mainly been at descriptive stage, introducing various factors that theoretically have shown effect on innovation systems.

Carneiro (2001) believes that since the innovativeness potential of the firms is dependent on the extent of knowledge diffusion among the groups of individual agents, which by itself is affected by the structure of the organization and network, in various studies, the focus has been on the social features of innovation networks. In social network analysis, the individual agents in various firms are the nodes and any relationships between any two of them in terms of knowledge exchange are considered as the links in the network (Carneiro 2001).

The innovation networks as social networks have been studied in two distinguished levels in the past literature. First is the level of innovators, who are the individual scientists from different organizations with a number of publications and patents, and are collaborating with each other to produce new knowledge (Zucker and Darby 1996; McMillan, Narin, and Deeds 2000; Gauvin 1995; Smith 2007; Sosa and Gero 2005; Keller 1991; Macdonald and Williams 1994; Macdonald and Williams 1993). The second is the inter-firm collaboration level that is a set of firms and companies involved in the research collaborative partnerships with other firms (Aharonson, Baum, and Feldman 2004; Powell, Koput, and Smith-Doerr 1996; Van De Ven and Rogers 1988; Damanpour

1987; Kimberly and Evanisko 1981; Damanpour and Evan 1984). The results of former studies show that in either form of the networks, the tendencies to continue collaboration with previous partners are high (Mayer and Argyres 2004; Dyer and Harbir Singh 1998; Kale, H. Singh, and Perlmutter 2000; Kim and Vonortas 2005). This means that any pairs of individual inventors or firms, who have already come up with innovation together, trust each other and are more prone to keep in touch for their further innovations, and their relationships may last even after one innovation is introduced.

Conclusion: The learning processes are faster on the innovation networks, since the exposure to various sources of knowledge is high. The social context of innovation networks has been studied as the main basis for the knowledge transmission. The literature targets the network of individuals and firms collaborating with each other as the nodes of the networks. In this level of study the social features of the agents are studied.

2.4 Simulation approach

Becheikh et al. (2006) reviewed empirical studies that try to find factors and variables determining the progress of innovation in manufacturing firms, from 1993 to 2003. They categorize these factors into two major groups: 1- inter-firm factors, 2- external factors.

The inter-firm factors, according to Becheikh et al. are: general characteristics of firms such as size, age, ownership, etc; cooperative and business strategies of the firms; firm’s structure; management structure; functional departments and their strategies.

On the other hand, Becheikh et al. consider the following factors as external factors that authors have studied through literature: industry; geographical location; structure of the network; level of the technology and knowledge; government and public policies; culture.

Different authors have studied the above mentioned factors (Becheikh and Su 2005; Zaheer and Bell 2005; C. Wu and Zeng 2009; Wang et al. 2008; Shih and Chang 2009; etc), and for each of these factors, there are a number of articles that either support or reject their positive effect on the innovative behaviour of the firms. The diversification of the results from various authors suggests that the innovative behaviour of the firms in the industries is derived from a combination of factors, identification of which is a complicated problem that requires a more systematic approach, rather than empirical studies. For example, Rogers (2004) develops a survey using empirical data to evaluate the effect of various factors on innovation. The analysis outcomes show that exporting activities can only improve innovativeness in small sized firms, and does not have a significant effect on big firms. Therefore, the various possible combinations of the two factors and the complementarily among them, exporting activities vs. firm size, may have contrasting results on the innovation networks. Many of these contrasting results can be seen through the work of authors same as Rogers, who merely use empirical data to form their analysis.

The dynamic behaviour of innovation networks has not been discussed much. The introduced factors must be analyzed in order to measure the magnitude of their impact on the innovation networks, in the presence of various combinations of all other factors. Taking into consideration all the possible factors and studying the behaviour of innovation networks thoroughly requires an enormous model to be established, which can imitate the procedure of innovation networks in different situations.

One of the widely used methods in this regard, mostly in recent works, for evaluating these factors is simulation. Simulation models are developed to analyze and measure the impact of positive and negative factors in the networks, to develop various possible combinations of factors, and to compare these combinations.

Albino, Carbonara, and Giannoccaro (2006) categorize the manner of discussion about innovation into three groups: conceptual, empirical, and simulation based. They suggested that now it is time to analyze the presented theories about innovation network's characteristics through simulation. As mentioned before, although the economic, sociological, management and policy based studies have clarified the role of innovation networks in the process of knowledge distribution, there are as yet many fundamental questions about the dynamics of the innovation networks.

Innovation networks and most of their features and behaviours are abstract concepts that need additional tools to be analyzed and understood effectively. Nowadays, computer-based software are helping researchers simplify the realization process of abstract systems and analyze various situations. Among all available computer software, simulation software is the most useful in analyzing the performance of the complicated abstract systems.

Since organizations and firms, as the agents of the innovation networks at issue, are autonomous units that act independently, and the changes in the whole network instead of its subsystems can show the evolvement of the networks over time, **Agent-Based Modeling (ABM)** simulation software is the best solutions for modeling such systems, and have been used for these kinds of problems in the previous works of several authors (Pyka, Gilbert, and Ahrweiler 2009; Gilbert 2004). According to Ma and Nakamori (2005) ABM is a very powerful tool for examining the "complex adaptive systems". They also believe ABM is generally beneficial in modeling the complications from micro-level activities on the macro level agents and systems. Pyka and Fagiolo (2005) state that ABM is also capable of modeling the related complex phenomena, their emergence, growth or destruction, and is a good tool for depicting various incidents in time, both quantitatively and qualitatively. ABM is said to allow the managers and decision makers to compare several scenarios of development.

Gilbert, Pyka, and Ahrweiler (2001) are among many authors that have studied innovation networks through simulation. They believe that the positive role of innovation networks has been examined, but the literature does not discuss the creation process of innovation networks, the procedure of its expansion, destruction or merging with other ones. Gilbert et al. have developed the simulation model of a general innovation network, in order to examine the outcomes of various situations and conditions, and also verify the assumptions and theories behind that. According to Gilbert et al. simulation uses assumptions to generate data on which further analysis and deduction can be performed. The developed ABM in their survey primarily tries to simulate the procedure of formation, growth and death of firms in the innovation networks. Gilbert et al. suggested that the simulation model can be used to examine the consequences of policies and changes and is a tool to help decision makers decide based on the possible results of the decisions. Their essay evaluates three case studies by means of simulation: telecommunication, biotechnology and electronic business. The results show that the initial distribution of agents is very much important on the formation and growth of the networks. There can be two main types of agents in this regard: large firms, and small firms. The extent of investment on innovative activities is also observed to play an important role. Besides, the firms that choose to cooperate with others are more likely to survive and grow than those trying to use only the inter-firm R&D. Gilbert et al. also concluded that if a network dissolves, the agents will have difficulty in connecting to other networks and are more prone to death. In addition, any firm that gains bigger amount of capital (reward) as a result of its activities in the initial stages is more likely to succeed further in the model. Finally, Gilbert et al. imply that there is still much work that can be done on simulating the innovation networks, and more detailed computer simulation models can be developed in the future in order to analyze the complicated behaviour of various innovation networks with different characteristics.

Albino et al. (2004) define industrial districts (IDs) as a group of firms that are engaged in manufacturing of similar products. In other words, industrial districts can be thought of as a special kind of geographical cluster, in which the firms are highly specialized in some aspects of the procedure. Specialization leads to the improvement of knowledge, and learning processes. Two major learning processes are suggested by Albino et al. (2004): traditional and new. Traditional learning process refers to the innovation that is achieved by using the distributed knowledge and information inside the IDs, i.e. learning based on the interaction of firms. On the other hand, new learning methods are innovation processes that use the knowledge from outside of the IDs, such as universities and research centers, or try to produce new knowledge by investing on R&D activities inside the IDs. Albino et al. introduced, developed and simulated four different scenarios to answer the questions of their survey: In which scenarios are IDs more innovative, when the demands of clients are rapidly changing? Which learning process should be adopted? And what is the role of the leader firms?

The two decision factors are probability of demand of new products in IDs, and the probability of high rate of required innovativeness, in two levels, low and high. An agent based simulation model is then developed to evaluate the impact of each learning method, and role of leader firms in industrial districts. Albino et al. concluded that in order to improve the innovative capabilities, firms inside the IDs must adjust their innovation processes to the needs of a highly competitive environment. The traditional learning processes should accompany the new ones in order to provide the needed knowledge and information to compete in today innovative world. The existence of leader firms is also investigated to be important for IDs to increase innovativeness, but is not sufficient by itself.

According to Tian and Zhang (2008), the formation and growth of innovation networks are highly dependent on the properties of the industrial clusters, in which they arise, and as the industrial clusters evolve overtime the innovation networks continuously grow as well. In the ABM simulation model developed by Tian and Zhang, the knowledge level of agents has been considered to affect the success of innovations. The distribution rate of new knowledge, called transferability, the direct or indirect relation between agents that use the distributed knowledge, and internal capability of agents to gather knowledge are also the factors considered in the model. Tian and Zhang assumed that there is a negative relation between the distance of two agents and the probability of one being selected as a partner. Tian and Zhang also assumed that agents are more likely to reconsider their old partners than starting a new relation. Making any connection is assumed to cost the agents the amount that is proportional to the geographical distance between agents. The results of Tian and Zhang's model imply that the innovation networks in industrial clusters tend to grow as the knowledge difference and specialization level of agents increase. And also the preferential rule of selecting partners and the increasing cost of new connections, as the distance increases, result in small networks that are geographically close inside, and have some connections to other further networks (small world effect).

Pyka et al. (2009) developed an ABM simulation study to test the hypothesis claiming that innovation networks are temporary structures that disappear after the organizations grow enough to act independently. Their results were different. Pyka et al. claim that the creation and development of the innovation networks happen in a recurring manner, which means that some of the small networks can grow enough to join the main network, and some others may disappear temporarily due to lack of strong connections between their firms. They also claim that the presence of the big firms at the initial stages of network development has a positive effect on the survival and growth of the networks. They observed that the more the firms/ scientists are loyal to their previous partners, the more it is likely for the whole network to grow and become scale-free, which confirms the earlier theoretic hypotheses regarding the impact of loyalty to previous partners introduced by Mayer and Argyres (2004), Dyer and Singh (1998), Kale, Singh, Perlmutter (2000), and others.

Gilbert et al. (2001) generated an ABM simulation model as well, in order to compare the situations in which firms and organizations are more willing to either invest on their own internal R&D, or create collaborative interactions with other ones. They, via simulation, show that in some situations and under some parameter settings of the model, which depends in the network being studied, investing in firm's own R&D produces better results, and with some other settings collaborative interactions are preferred. The fact that various settings of the model result in different answers also supports the idea that by creating more specified models of target networks, more precise conclusions can be drawn for the policy and decision makers (Gilbert, Pyka, and Ahrweiler 2001).

There are yet some other simulation based studies that have been carried out to clarify the current hypotheses about the innovation networks (Udell, Bottin, and Glass 1993; Guardiola et al. 2002; Gilbert, Ahrweiler, and Pyka 2007; Wang et al. 2008; Wu and Zeng 2009; Lenz-Cesar and Heshmati 2010).

Conclusion: There are three methods for analyzing any type of system: direct experiment, mathematical analysis, and simulation modeling. According to the literature, simulation modeling is an applicable approach for studying the dynamics and characteristics of the innovation networks. Most of the literature in this regard is targeting the firm-level of study, and less are devoted to the individual-level of innovation networks. Besides, there is no simulation approach in the literature using the realistic data, and the simulation models are built up according to the hypotheses.

2.5 Gatekeepers and Star Scientists

As described before, the information transmission, which results in novelty and creativity, is the key characteristic of the innovation networks and clusters. The information transmission occurs through formal or informal channels, inside or between organizations, which are established based on the personal contacts of individuals. Of course not all of the transmissions lead to novelty, and not all of the individuals are innovators. However, the individuals who are formally responsible for providing the channels and link separate sources of knowledge are defined as Brokers (Marsden 1982).

The role of the intermediaries in the innovation process as the facilitators of transition of knowledge and information has attracted a lot of research interest. The most important role of brokers is to make connections between actors, who do not have either approach to or trust in each other (Marsden 1982, 202). Not all the firms are connected to each other in a network, and therefore there are many “structural holes” in the network preventing the information from flowing evenly among the firms.

Burt (1992) defines a “structural hole” as a disruption in the flow of information between agents in a large network, as a matter of lack of connection. The more successful a firm is in filling up these structural holes, the greater advantage it has over competitors. Burt (1980) and later Fernandez and Gould (1994) argued that organizations with brokerage position in the networks are more influential than others. However, according to Burt (2007), only the immediate network around a broker is beneficial, and all indirect links for the broker in the network have significantly reduced profit.

The intermediary role of transferring knowledge between sources of new ideas in the innovation networks is not all a broker does in the network. Since brokers transfer knowledge, they are also valuable depositories of knowledge, who are able to merge different existing ideas and come to new solutions in various industries (Hargadon and Sutton 1997). Among all the roles suggested for brokers[4] (Gould and Fernandez 1989), the act as a Gatekeeper is specially addressed in the present thesis.

The information can be categorized as related and unrelated. The related information is easily detectable and the information systems inside organizations can search for them and provide them to the demanding parts without any problem. Though, most of the times innovation occur when some seemingly unrelated and unusual information is provided, and new solutions and creative ideas are brought on. The role of the Gatekeepers consists of the recognition and gathering of this unusual information at the right time, and more importantly, providing it to the right sub-system in the organization (Sosa and Gero 2005).

The term "Gatekeeper" has been used in various related articles with meanings slightly different from each other (Sosa and Gero 2005; Keller 1991; Macdonald and Williams 1994; Morrison, Rabellotti, and Zirulia 2008; Smith 2007). In the present thesis, only

[4] According to Gould and Fernandez (1989), the mentioned roles for the brokers in the networks are:
Gatekeeper: the one who decides to grant access to an outsider
Representative: assigned member of a subgroup for communicating with outsiders
Liaison: a link between two subgroups without commitment to them

those individuals who are primarily inventors by themselves, and make connections between two or more separate clusters by bridging their information flow, are called Gatekeepers.

According to Schiffauerova and Beaudry (2008), usually only a maximum of one fifth of the innovators in the networks are accountable for absorbance of external fresh knowledge in a cluster. The creation of connections between Gatekeepers takes time, trust and money for both parties. The low number of the Gatekeepers, as well as their crucial role in transmitting the confidential or general knowledge makes the organizations concerned about the sustainment of the Gatekeepers within their positions (Macdonald and Williams 1994).

The question that arises is what is going to happen to the connection, and consequently to the network, if a Gatekeeper is no more in its position, or lost. The related literature regarding the Gatekeepers and their role in the networks has seldom considered this question. Sosa and Gero (2005) point out that most of the previous work around creativity and innovation think of it on the basis of individualistic assumptions, which look at the innovativeness as an extraordinary ability of a single person. Though, innovation is usually a novel solution given to an existing problem, and most of the times generated after a group of people are engaged in the problem and various possible solutions are brainstormed. Therefore, Sosa and Gero (2005) define Gatekeeper as an opinion leader who manages the process of innovation by controlling the selection, feedback and assessment of the new ideas.

Obviously, this definition is different from the view point of the present thesis, in that it does not consider the Gatekeeper's knowledge transfer characteristic. However, Sosa and Gero's findings related to the common characteristics of the innovation networks and Gatekeepers are interesting for this work, since they have also developed a simulation model.

Sosa and Gero (2005) believe that the social ties are playing an important role in controlling the power of Gatekeepers in the networks in the way that in the networks with strong social ties between Gatekeepers, "lower mobility and hierarchical structures of influence exist". They conclude that in societies with strong ties[5], gate-keeping is more unchangeable in a way that only a small group of experts are always playing the Gatekeeper role, while in weakly-tied societies Gatekeeper role is rather distributed among the agents and does not represent a consistent behaviour. It also can be concluded that the ties and links can determine the effective power of Gatekeepers, and consequently the sensitivity of the network to their presence.

On the contrary, Macdonald and Williams (1993) did not see any necessity for a Gatekeeper to be an expert. They define Gatekeeper as a know-who instead of a know-how, in sense that they "know who knows" the particular information outside of their organizations. They believe that Gatekeepers have enough knowledge to use the information they gather, and even better, they can decide which information is worth gathering and which is not.

---

[5] "Strong social ties usually exist between nodes in a kinship network, while weak ties characterize networks where casual encounters occur between strangers or acquaintances" (Sosa and Gero 2005)

In the age of change and flexibility, and where most of the telecommunications are formal, it is vital for the organizations to be alert to new information as a whole, instead of only some individuals, Macdonald and Williams argued. However, the role of Gatekeepers in transferring the information that triggers innovation is still as important as before, they believe. Macdonald and Williams also distinguish Gatekeeping from the act of mere information bringing. In their opinion Gatekeepers still coexist with other methods of communication.

Obviously, the more links a company or scientist has to outside sources of knowledge, the more amount of fresh and new knowledge it can access and bring into the cluster for further innovative activities. However, the big number of links by itself does not guarantee the quality of innovations, and the capabilities and experience of research personnel and scientists are also key factors in determining the excellence of results. The ability of a scientist in detecting useful knowledge and combining it with already existing one in order to produce innovative results highly depends on his previous research activities and background. Consequently, the productivity rate of innovativeness in a firm is both affected by the number of its external links as well as the research quality of its scientists. (Henderson and Cockburn 1996)

Zucker and Darby first in 1995 and then in 1996, applied the term “star scientist” to qualified research personnel of firms, who improve internal research productivity by their excessive experience in research and innovative activities. In the other words, star scientists are researchers with significantly higher productivity in terms of innovation and knowledge development than their colleagues and rivals.

Although the qualitative definition of star scientists as researchers with greater number of discoveries, articles and patents than other equivalent researchers is clearly understood and accepted at the first sight (Zucker and Darby 1996; Queenton and Niosi 2003; Cohen and Levinthal 1990), the exact qualitative definition of them varies from study to study based on the characteristics of the networks, their organization types and major of activity.

The benefit of Gatekeepers and star scientists for networks is the improvers of knowledge circulation and new knowledge generators respectively, as the former bring new fresh knowledge to the cluster (Schiffauerova and Beaudry 2008), and the latter have enough experience to utilize this new knowledge and produce innovative outcomes out of it (Smith 2007; Keller 1991; Zucker and Darby 1995).

In the literature, there is not much said about the characteristics of the Gatekeepers and star scientists. There are only various studies about their behavioural effect in the aspect of psychology and society, which seem unrelated to the purpose of this thesis. One of the closely related studies about the role of Gatekeepers has been done by Keller (1991), who introduces the following characteristics based on an empirical study in U.S. and Mexican organizations: concentration and proximity of Gatekeepers in strong organizations; higher performance that usual employees i.e. higher number of patents and publications; similarity of action in various industries.

Overall, there is a lack of research on the formation of innovation networks in the presence and absence of the Gatekeepers and star scientists, and still more study is needed to enable us to understand, analyze, and predict the alteration of innovation networks in various possible combinations of their components, i.e. scientists, Gatekeepers and stars, in addition to the above mentioned empirical and tentative studies. Therefore, this thesis poses its two main research questions as: What is going to happen to the innovation networks, if a Gatekeeper or star scientist is lost? One of the objectives of the present thesis is to provide answers to these questions.

Conclusion: The literature categorizes the role of scientists in the innovation networks according to their affection on the creation and circulation of the knowledge. The main creators of new knowledge are called stars, and the intermediary scientists responsible for the inflow of knowledge to the cluster are called Gatekeepers. There is a vast discussion of the role of the stars and Gatekeepers, and their characteristics. However, almost all the surveys are qualitative, and there is no quantitative definition or examination on the role of different types of scientists in the networks.

2.6 Canadian Biotechnology Networks

Biotechnology is a sub-field of biology science that combines biology with engineering, chemistry, physics, computer science and mathematics, in order to utilize and modify living organisms according the human needs, and develop bio-products (Shmaefsky 2006). Biotechnology is said to be a very young science and fast growing compared to other fields of science. Since it was first introduced in 1919 by Ereky, biotechnology has found many applications in various fields of science, such as agriculture, energy production, environmental sciences, manufacturing, and medicine all over the world (Fári, Bud, and Kralovánszky 2001).

The Canadian biotechnology society (Biotech Canada)[6] defines biotechnology as "applied scientific disciplines including chemistry, engineering, physics, and computing to living organisms to create innovative products and techniques". After 1980, Canada has become the second pioneer country in Biotechnology, after United States, mainly based on its number of firms, patents and innovative products (Niosi and Bas 2004).

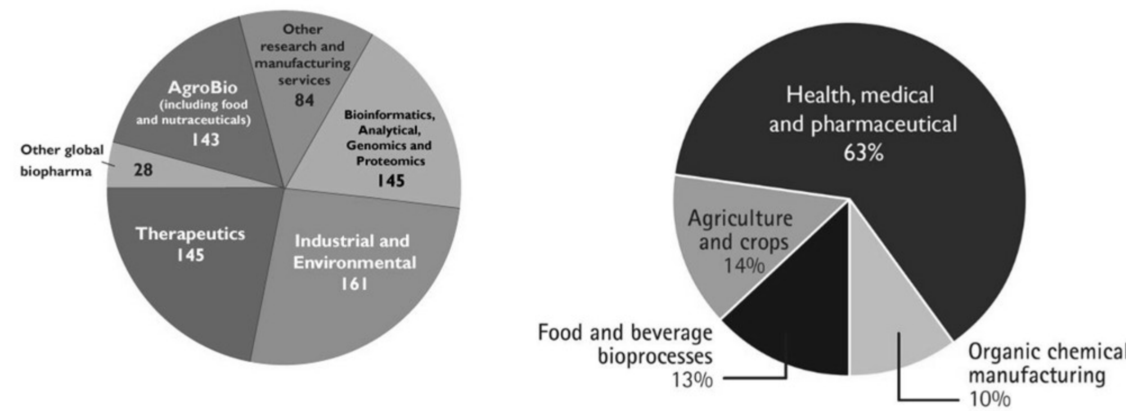

Figure 1: (Left) Canadian Biotech Companies' Working Field, (Right) Canadian Bioeconomy Components Distribution

Canadian biotechnology has three main categories: health, agriculture, industrial. The largest sector of Canadian biotechnology is health, consisting of four subcategories, Therapeutics, Diagnostics, Medical Devices, and Vaccines. According to the latest statistics published by BioteCanada[6], there are a total of 583 Canadian biotechnology companies. Figure 1 (Left) shows the working fields of these 583 companies. Among these companies, 35% belong to western Canada (Alberta and British Columbia), 28% to Ontario, 21% to Quebec, and 16% to other regions. With 7.0 percent of the GDP,

[6] http://www.biotech.ca

biotechnology constructs the third economical fundament in Canada. According to Statistics Canada[7], the distribution of main components of Canadian Bio-economy is as presented in Figure 1 (Right).

Canadian biotechnology innovation networks have recently attracted the attention of some network researchers. J. Niosi and Bas 2001 study Canadian innovation systems and believe that as a knowledge-intensive research field, biotechnology clusters are prone to geographical concentrations because of the high amount of spillover in them. They study the regional partitions of Canadian biotechnology, and categorize Toronto as the leader cluster in Canada's biotechnology. After that, Montreal and Vancouver are ranked as the second and third clusters. They also show that spillover, between Canadian biotechnology clusters, is dependent on region, organization type, and also size and characteristics of the market such as availability of capital.

Niosi and Bas (2001) categorize universities and government laboratories as first and second organizational sections of biotechnology in Canada. They show that "star scientists" as the most productive innovators in Canadian biotechnology network are mainly working at the universities of the large cities in Canada. Therefore, according to (Niosi and Bas 2001), universities handle the main research activities in Canadian innovation system.

Niosi and Banik (2005) suggest that in Canadian biotechnology innovation network, the three largest provinces, Ontario, Quebec, and British Columbia, are the center of innovation, and as the distance from them grows, the number of innovative firms

[7] http://www.statcan.gc.ca (2011)

decreases. Therefore, the concept of location seems to play an important role in Canadian innovation systems. Besides, they claim that more patent licenses are granted in these provinces than other regions.

Schiffauerova and Beaudry (2011) also have studied the structural characteristics of the Canadian biotechnology network and the role of star scientists on the Canadian biotechnology network and showed that stars have more collaborators and have better access to the flow of knowledge in the network. They also use empirical data about patents and patent ownership, and number of citations to prove that most of the star scientists and more than half of the highly cited scientists are also Gatekeepers in Canadian biotechnology networks.

As for conducting an empirical analysis of the innovation networks, and more precisely for developing a simulation model for this purpose, the primary necessity is having real data of an already established network. Most of the empirical analyses conducted by other authors also are based on case studies of various clusters and networks around the world (Acs and Audretsch 1988; Beaucage and Beaudry 2006; Gemünden, Ritter, and Heydebreck 1996; Keller 1991; McMillan, Narin, and Deeds 2000; Schartinger, Schibany, and Gassler 2001).

According to Gemünden, Ritter, and Heydebreck (1996), at the time there were two general types of empirical analysis in the literature regarding the interactions in technological environments that may result in innovative accomplishments: studying the relationship between crucial scientists by conducting face-to-face interviews; quantitative and large-scale analyses through mailed questionnaires.

However, since 1996 the data collection technologies and methods have improved significantly, and nowadays computer-based databases are available including the details of scientists' performances in the networks. As one of the very primary steps in simulation modeling, data collection is then highly dependent on the case under study, of which the crucial and specific characteristics of the networks could be detected and applied to the model (Velten 2009).

The main focus of the present thesis is on the Canadian Biotechnology patent and scientists, the data for which is extracted from USPTO (United States patent and trademark office) and SCOPUS[8] respectively. The data consists of all the biotechnology articles and patents with at least one Canadian author/ inventor. USPTO is the only patent database that includes the geographical location of residence of each inventor. So, the combination of databases helps to have a realistic analysis of the innovation network among Canadian biotechnology scientists.

As a similar study to the present thesis about the Canadian biotechnology innovation network, (Aharonson, Baum, and Feldman 2004) study the extent of innovativeness in firm level of the network, constructing four empirical analyses on a US patent database. Considering the Canadian biotechnology industry in their paper, they investigate the level of influence that the R&D activities of one's own firm and other geographically local firms may have on the firm's innovative output, which is estimated based on the patent application rate of the firms.

---

[8] SciVerse Scopus is the world's largest abstract and citation database of peer-reviewed literature and quality web sources. (http://www.scopus.com)

Aharonson, Baum, and Feldman (2004) suggest that using realistic data that is confined to a heterogeneous industry can specify a more detailed model that excludes uncontrolled characteristics of firms. Moreover, using a highly detailed database provides them with much more specialized information about knowledge spillover among firms of the same kind. This way, stronger conclusions can be made about spillover effects than previous related studies, which hypothesised cross-sectional spillover among firms on much broader industries, without any actual proof gained from the real world.

Rampersad, Quester, and Troshani (2010) also believe that although there exists a vast literature on the principles of innovation networks and their characteristics, a few are dedicated to firm level and individual level of specific industries. They insist that without specialized information of a certain industry, decision making processes cannot be conducted in managerial levels of study, i.e. new product development and marketing. Therefore, they have established their studies based on the information from high technology networks, including biotechnology. According to them, there is a gap in the literature for exploring diverse network actors such as universities, research organizations, government research groups, etc.

In addition, Gittelman (2007) believes that learning the principles of innovative communities in order to extract regulations for them requires a more detailed picture of each society, so that they can benefit from specialized approaches appropriate for their kind of activity.

As a result, in order to investigate the innovation networks more precisely, and develop applicable guidelines and instructions for various structures of the networks, it seems necessary to analyze the systems based on the information extracted from recent activities in them. In support of this notion, and with the intention of understanding the structure of the Canadian biotechnology innovation networks, the simulation model built for this thesis, as mentioned before, uses realistic data about patents and publications of an individual-level perspective of scientists in Canada from 1952 to 2006.

2.7 Conclusion

Innovation is said to be the main competitive advantage of firms. It emerges in the form of any change in the existing thoughts or procedures, and aims at improving the human life (Gopalakrishnan and Damanpour 1997). The development process of innovation has therefore attracted attention in various industries. Knowledge is known to be the key factor for innovation, and knowledge transfer mechanisms have been known as the main basis for innovative activities (Marshall 1961; Porter 1998; Chuanrong Wu and Deming Zeng 2009; Wu and Zeng 2009).

The concept of innovation networks has attracted attention recently, and in the late twentieth century the role of intermediaries, clustering and networking in the distribution and improvement of knowledge was almost clarified (Howells 2006). Various determinants and characteristics have been identified for the innovation networks, and different hypotheses have also been introduced, among which geographical position of the agents in the network and the characteristics of individual scientists have been discussed the most and the least respectively.

As it is stated in the literature, scientists are more interested in the collaboration with people in their region, rather than people in long distances or overseas (Schiffauerova and

Beaudry 2008; Fleming, King, and Juda 2007; Martin and Sunley 2003; Gittelman 2007). The concept of clustering, as geographical concentration of firms and organizations with purpose of exploiting available knowledge spillover from competitors, has been in the center of attention in analyses of innovation networks. (Baptista and Swann 1998; Baptista and Swann 1999; Porter 1998; Martin and Sunley 2003; Prevezer 1997; Swann and Prevezer 1996)

Beside geographical distance, the influences of other factors such as organizational size, level of specialization, loyalty to previous collaborators, and inter firm activities in the innovation networks have also been studied (Plank and Newell 2007; Damanpour and Evan 1984; Kimberly and Evanisko 1981; Damanpour 1987). Furthermore, the impact of individual scientists on the transmission and improvement of knowledge in the networks has been analysed (Schiffauerova and Beaudry 2011; Zucker and Darby 1996; Macdonald and Williams 1993; Zucker and Darby 1995).

Understanding the dynamics of innovation networks can help in improving policies and rules for firms and governments and lead to the enhancement of innovation, both quantitatively and qualitatively. In order to understand the dynamics of the observed characteristics of the innovation networks, and to study the magnitudes of the factors in various combinations, a systematic approach should be applied (Wu and Zeng 2009).

Methodologies employed for exploring different aspects of innovation networks vary from personal questionnaires to complicated simulation modeling and systematic analyses of the relationships in the networks. Among all the introduced methodologies, agent based simulation modeling of the networks seems to be a suitable methodology for examining their dynamics, since it provides a vast and detailed picture of the activities among individuals, firms and clusters in the network (Gilbert 2004).

However, the complexity of the innovation networks and the detail and the extent of information required for developing the simulation models have acted as a drawback for using this approach more frequently, and there are thus not many instances in the literature studying the dynamics of the innovation networks through simulation. The few instances that exist also do not cover all the many aspects of networks together, probably since such an inclusive modeling requires a highly detailed data gathering and powerful software. Most of the simulation models developed for analysis of innovation networks are agent based model, which are said to be the most appropriate tool for imitating the complex, multi-agent behaviour of innovation networks (Lenz-Cesar and Heshmati 2010; Albino, Carbonara, and Giannoccaro 2006; Ma and Nakamori 2005).

The role of intermediary individuals has been classified in this thesis into two major groups, star scientists and Gatekeepers. Stars are the most prolific scientists in the network, generating more amount of knowledge than others (Niosi and Banik 2005). Gatekeepers are individuals responsible for bringing fresh knowledge to their clusters by bridging over long distances (Macdonald and Williams 1994). The basic characteristics of stars and Gatekeepers are introduced in the literature. However, no research study so far has focused on the dynamic performance of networks in the presence and absence of stars and Gatekeepers.

Finally, innovation network of biotechnology, as an important and fundamental technology that feeds many other industries and technologies, has been the subject of recent research on the topic of networks (Schiffauerova and Beaudry 2011; Rampersad, Quester, and Troshani 2010; Schiffauerova 2009; Schiffauerova and Beaudry 2008; Gittelman 2007; Beaucage and Beaudry 2006). Understanding the performance of knowledge distribution and its stimulators can provide a vast understanding about the networks and trigger a big step toward further innovative accomplishments in biotechnology and related industries. The performance of Canadian biotechnology network has been analysed theoretically and the activities of stars and Gatekeepers have been investigated (Queenton and Niosi 2003; Niosi and Bas 2004; Schiffauerova and Beaudry 2011). However, simulation-based studies are needed to understand the dynamics of Canadian biotechnology network, in order to improve its performance.

3. Research Questions and Objectives

3.1 Research Questions

Since the main focus of the thesis is on the interactions of scientists in the Canadian innovation networks, the innovative productivity of the scientists shape the heart of the thesis. The relationship between the productivity of the scientists and the characteristics of their network are studied in this thesis.

First, the structural behaviour of the network is investigated to study the behaviour of various factors, such as geographical clustering, scientist's background, and organization types on the innovative rate of the Canadian biotechnology network.

Second, the impact of the loyalty is investigated in the networks. The consequences of loyalty to previous partners are determined by Mayer and Argyres (2004), Dyer and Singh (1998), and Kale, Singh, and Perlmutter (2000), suggesting that scientists remain loyal to their previous partners and prefer to work with them again in their future works, which can affect the whole network in a way that beginner scientists have lower opportunity for future works, and may be disconnected from main sources of innovation flows and become isolated. However, the effects of the loyalty on the structure of innovation networks has so far not been analysed quantitatively. In the present thesis, the impact of loyalty is analysed in Canadian biotechnology networks in order to shed more light on this topic and to understand whether there are advantages or disadvantages for the repetitive relationships in terms of the overall innovation efficiency of the network.

Third, the presence of star scientists is investigated, and the effect of their activities and their presence on the network structure of Canadian biotechnology clusters is also studied. At the theoretical level, the impact of star scientists on the growth of innovation has been studied in the literature (Schiffauerova and Beaudry 2011; Zucker and Darby 1996; Niosi 2006; Smith 2007; Macdonald and Williams 1994; Keller 1991). In the present thesis, their roles are analysed in the Canadian biotechnology network with empirical data to compare the performance of the networks in their presence and absence.

Finally, the role of Gatekeepers is investigated in the network. Gatekeepers have been shown to play a critical role in the flow of knowledge between geographical regions in the networks (Macdonald and Williams 1994; Macdonald and Williams 1993; Keller 1991; Schiffauerova and Beaudry 2008). In the present thesis, the consequences of their presence and absence in the Canadian biotechnology network are investigated.

3.2 Research Objectives

Objective 1: Simulate a general innovation network

- Develop a computer based model imitating the relationships and characteristics of innovation networks in scientist-level of study
- Make the model as flexible as possible for future investigations
- Determine the variables that could be included in the scientist-level network from the database of affiliations

Objective 2: Examine the loyalty of the scientists toward previous partners and its impact on the innovativeness

- Determine the performance of network with and without the loyalty factor
- Analyse the growth of network in case of high loyalty levels

Objective 3: Study the role of Gatekeepers in Canadian biotechnology network

Describe the performance of the model in the presence and absence of Gatekeepers

Determine the impact of Gatekeepers on the innovativeness in the network

Objective 4: Study the role of star scientists in Canadian biotechnology network

Describe the performance of the network in the presence and absence of star scientists

Determine the innovativeness in the absence of star scientists

Determine the impact of star scientists on the innovativeness of networks

4. Methodology

4.1 Innovation networks

The innovation networks considered in this thesis are formed by a group of individual scientists which are characterized by various real-life attributes. These scientists collaborate with each other in order to exchange knowledge and information, and thereby produce new scientific knowledge leading to an innovative output in the network. The scientists are represented by the nodes of the network, while the collaborations between them are the links connecting these nodes. In the real life the scientists meet for many possible reasons, but not all the connections between two scientists necessarily lead to the creation of new knowledge. In this thesis, the main concern is about the links that help improve the total level of innovation in the network, i.e. only the connections which resulted in the publication of a scientific article or the registration of a patent are considered. The innovation networks are thus built through mapping the article co-authorship and the patent co-inventorship linkages (the citations are not included). Such networks represent the creation of the scientific knowledge and of the innovative outcome through a complex web of knowledge-based relationships.

4.2 Data - USPTO and SCOPUS Databases

In order to create the networks an extensive data was needed. Two databases of the publications and patents[9] in Canadian biotechnology are available. The main approach of the thesis consists of the exploitation of the large amount of information extracted from the patent database of the United States Patent and Trademark office (USPTO)[10], and SCOPUS, which is the abstract and citation database of peer reviewed research literature. The information gained from the data is used to establish a simulation model of the Canadian biotechnology network and analyse the dynamics of this network.

The required information is extracted from the database, using an automated extraction program. The necessary information includes patent name, and inventor's name and address, for all the Canadian patents registered before March 31, 2007 (Schiffauerova 2009).

USPTO is the only patent database which provides the geographical location of the residence for each inventor. In the USPTO available patent database, there is a total of 104792 distinct patents, each with its own name, inventor's name, application number, year of application, year of granting, and number of claims. In the SCOPUS database, the database of biotechnology articles from 1952 to 2006, there is a total of 100750 articles,

---

[9] According to USPTO official website, a patent is defined as "a property right granted by the Government of the United States of America to an inventor "to exclude others from making, using, offering for sale, or selling the invention throughout the United States or importing the invention into the United States" for a limited time in exchange for public disclosure of the invention when the patent is granted."

[10] The patent database available for the empirical study of the biotechnology clusters is the United States Patents and Trademarks Office (USPTO) database since it is the only patent database containing the geographical location of the residence for each inventor. The use of the USPTO database may introduce a bias in the data, which is expected to be minimal, because Canadian inventors usually patent both in Canada and in the US. Canadian biotechnology firms prefer to protect their intellectual property in the USA, since the much larger and easily accessible US biotechnology market offers great potential to Canadian biotechnology firms. An analysis of the Canadian patents registered at the USPTO should hence provide a realistic picture of Canadian biotechnology innovation.

each with the complete information on the name of the article and the journal where it was published, year of publication, number of references, authors' first and last names, country, affiliation for each co-author and cluster into which each co-author has been assigned.

Less than 2000 entries in the two databases have one or more blanks in the records, which are omitted in the analysis process in order to have a homogenous input. Since 2000 defective entries out of more than 100000 represent less than 2% of the whole data, and also considering the fact that the entries are independent, the omission does not affect the investigation.

As long as the main purpose of modeling the Canadian biotechnology network in this thesis is to represent a true picture of real incidents in that network, only complete and appropriate data are considered for the simulation phase. The two databases are then analysed and necessary queries are created in order to calculate the input formula and parameters for the model, which is discussed in detail in input analysis part of the thesis. The examination of the Canadian patents and publications at USPTO and SCOPUS would help to attain a more practical demonstration of Canadian Biotechnology innovation network for further investigations of the components and procedures.

4.3 Simulation modeling

The aim of this thesis is to study the dynamics of the innovation networks. For this purpose, the most suitable methodology was determined to be the simulation of these networks. Simulation modeling, according to Banks and Carson (1984) is one of the operational methods for studying a system, especially abstract or unavailable systems, for which implementations of direct experiment or mathematical analysis are not possible. Simulation modeling is mainly used for analyzing dynamic and complex systems that evolve over time.

The system of collaborative scientists is a dynamic system of individuals with certain links among them at different points in time. As a dynamic network, Canadian biotechnology network shows a changing behaviour over time in terms of number of scientists, their collaborations, and the innovative products. In order to capture the performance of the network and analyse its characteristics, a time-based approach is needed to depict the evolution of the network over time.

Generally, simulation modeling refers to any re-creation of a real or abstract system, in order to examine its behaviour and characteristics over time, without directly interacting with the real system (Rossetti 2010). It is mainly a combination of system modeling, computer programming, design, engineering, probability and statistics. Basically, as a system evolves over time, its alterations and behaviour can be modeled via simulation. For this purpose, the principles of the real system should be created and validated in the form of mathematical and logical relations between agents of the system (Banks and Carson 1984).

Developing a simulation model enables researchers to answer to "what if" questions about the system, and lets them examine various possible scenarios and their impact on the system, without actually running the scenarios on the system. Simulation is therefore a practical tool for analyzing the performance of both already existing and to-be-developed systems and designs. It is also very useful in decision making situations for defining new policies, rules and procedures, since it helps managers observe the results of their decisions on the whole system before actually setting them (Rossetti 2010).

As the main target of this thesis, the network of scientists in Canadian biotechnology field is an extremely large and expanded network. The observation and examination of the components and their behaviour over time in such a vast network requires a systematic approach. There are hypothesised characteristics for various types of components (star scientists, Gatekeepers, etc) in innovation networks, that remain at a theory level in the literature because of lack of access to all the components. However, a simulation model of such a network can assist in analysis of various hypotheses without directly conducting time consuming and expensive experiments on it.

Today, technology has made simulation modeling a lot easier with the aid of computer programs and software that enables scientists to create even the most detailed simulation models of the real-world or abstract complex systems, and observe their evolution over time. One of the most important key advantages of simulation modeling in comparison with other methodologies is its representational ability in demonstrating the complex interrelationships and interactions between the agents of the systems. This ability is especially of value in this thesis, and can exactly lead to a better understanding of the Canadian biotechnology network and its characteristics, by signifying the collaborative interactions among innovative scientists. Nowadays, as the advantages of simulation modeling have become more and more understood, it has found new applications in many different areas such as manufacturing, logistics, military, social sciences, geology, telecommunications, etc. (Velten 2009).

In addition to its advantages, simulation modeling requires a vast amount of data and time to be established and despite the fact that it sounds to be a great tool for analyzing any system at first, simulation modeling is not a recommended technique until other faster and cheaper methods can be applied to the situation (Rossetti 2010). A faster and cheaper method for analysing the innovation networks could be sampling, which is specially utilized when some theories and principles about some certain aspects of a network are needed to be testified. However, sampling is not appropriate for studying the whole network, since not all the connections among the agents in a network could be captured, and sampling usually requires neglecting some of the links in order to limit the sample size and make it possible to study all its components individually (Hanneman and Riddle 2005). Although empirical analyses provide more accurate results which are based on direct examinations of the system, they are less applicable when the size of the system grows.

The most time consuming part of simulation modeling however, involves gathering information and data from the real system, and building up general mathematical expressions in order to imitate its behaviour correctly and precisely. Data gathering phase of simulation is therefore the most expensive phase of modeling as well, and depends on the availability of the system and its characteristics and complexity. Since simulation modeling saves researchers and managers from doing expensive or sometimes damaging experiments on the real system, there is a trade-off between the expenses of doing the experiment directly and creating the simulation model instead.

There is also another trade-off between the time of creating a model of the system and the time it takes to use other solutions for the analysis process. Considering the advantages and disadvantages of simulation, and the fact that a simulation model is only useful and helpful when it is accurately built with sufficient data, it must be considered as a solution to the situation when all other methods are rejected (Banks and Carson 1984).

For the present thesis, the data gathering phase had been already done and the required data is available from the sources described before. This fact plays an important role when choosing the methodology for the thesis, letting consider simulation modeling as a way to find solutions for the research questions. The advantage of having the information of the whole Canadian biotechnology network simplifies the step of data gathering, and saves a lot of time and effort in the analysis process of Canadian biotechnology network. The availability of the data also rejected the need for implementing direct experiments in the network, and sampling from the network.

In the present thesis, the complex nature of innovation networks and relationships between the agents of the network, and also the previous similar works in this regard that have chosen simulation modeling as a clue for investigating the innovation networks, was a motivation to employ this approach as the methodology for the purposes of this thesis, and answering to the research questions of the thesis. The data gathering phase is already done by extracting required information from USPTO (United States patent and trademark office) and SCOPUS. The required information for each part of the thesis has been extracted from two complete databases available, defining appropriate Queries by SQL in combination with MS Access 2007. For building up the model, Java Developer language in accordance with Eclipse coding environment has been chosen. Although there already some pre-developed software for simulating agent based networks exists, it is chosen to develop the whole simulation model from its basic principles by author, so that the model could be completely flexible and open source for further similar analyses.

In the following sections, the created model is described in detail, and the interactions with databases are clarified.

4.4 Model Description and Input Analysis

The simulation model type selected for this thesis is an agent-based simulation model (ABM), since individual agents of the model are typically characterized as rational (personal reasoning behaviour), presumed to be acting in what they perceive as their own interests using certain decision-making rules. The system state variables (i.e. number of scientists, number of publications, and number of links) of multiple agents change at discrete points in time.

In the innovation network created for this thesis, the level of study is based on the scientists; in other words, scientists are the nodes of the network. Since the available data about scientists is on a yearly basis in the two available databases, and the time between publications is assumed to be on a monthly basis, the time slot for the model has been chosen to proceed monthly.

Based on the available data, the model is defined as follows: At the beginning of each year, a number of new scientists are added to the network, the trend for which is gained directly from the publications database (SCOPUS), by calculating the total number of new comers per each year. The yearly basis for new comers is mainly because of the structure of data in the two databases, but since the concept of time in the whole model is flexible (i.e. one can easily imagine that it is based on months, and then the results would be interpreted proportionally), this hypothesis for scientists to only enter the network at the beginning of each year does not affect the simulation results. In other words, since the evolution of the network is not affected by the time basis, and final number of links, nodes, patents and publications does not depend on the run-time clock of the model, this time frame is suitable for the purposes of the present thesis. However, other time bases can be used for different databases in the future.

For each new scientist, the new set of attributes and variables are defined with primary values. This set of attributes includes: scientist ID, geographical cluster ID, entrance year, organization type ID, research field, chooser threshold. These numbers are constant during the life of the scientist in the network. More details on the values of geographical cluster, organization type, research field, and chooser threshold are provided later in the present section of the thesis.

The set of variables for each scientist contains: Idle/busy status, number of articles, patent quality, journal's impact factor, number of links inside the cluster, number of links outside the cluster, score, age, star status, and Gatekeeper status. At the beginning, the idle/busy status is set as idle for each new comer. There is an initial value randomized uniformly for the number of articles, patent quality, journal impact factor, and the number of links inside and outside the cluster. This initial value is required for the programming purposes and does not affect the results of the modeling.

The star status of each scientist is set as true as soon as they meet some minimum specifications (i.e. the total number of their articles and/or patent quality is more than a flexible threshold defined in the model). Besides, as soon as a scientist has at least five links inside and two links outside the cluster, his/her Gatekeeper status is set as true.

The score is a number that evaluates the scientist's importance on the network based on its number of publications and patents and journal impact factor, and is mainly used in the choosing procedure of new partners for programming purposes. This number is calculated by multiplying the number of articles, patent quality and journal impact factor for each scientist. The summary of simulation modeling main parameters is presented in Table 1.

Table 1: List of simulation modeling parameters

Simulation modeling parameters		
Set of attributes	**Set of variables**	**Partner selection decisive factors**
Scientist ID Geographical cluster ID Entrance year Organization type ID Research field Chooser threshold	Idle/busy status Number of articles Patent quality Journal's impact factor Number of links inside the cluster Number of links outside the cluster Score Age in the network Star status Gatekeeper status	Geographical distance Organization type Previous collaborations Score Field of study Star status Gatekeeper status

At each point in the time slot of the model, first for each idle scientist in the network a random number is produced to decide whether they want to go to a busy status or not (more details on this number is provided in the "Idle/ busy status" section of the thesis). Then, all the scientists who are chosen to go to the busy status enter a loop. In the loop, a scientist is chosen and for him/her a random number as the number of partners is set. Then all other remaining scientists are searched to find enough number of partners for the present scientist. The scientists are selected as soon as the *chooser* variable meets the maximum of chooser threshold for both scientists. When all the scientists are assigned to a group, all their statuses are set as busy, and for each group a publication duration is set, during which all the scientists are busy and imagined to be working on the new publication (more details are given in the “publication duration” section).

For the scientists whose busy period is over, their number of articles, patent quality, and journal impact factors are updated. Then, a programming decisive variable decides whether the scientist remains in the network or dies. In the model, a scientist dies if one or more of the three death conditions are met: 1- his/her number of active links in the network fall to zero; 2- his/her age in the network passes twenty years (i.e. the longest observed life of one scientist in the two databases); 3- he/she is idle for more than 10 years. All these numbers are changeable in the model in order to analyse the behaviour of the network with various scenarios.

In order to picture the network of scientists, some information about the links between each pair of scientists is also required to be recorded. This information includes the age of the link and the IDs of scientists at each end. The age of the link is set to zero whenever two scientists stop working together. There is a link age defined in the model that causes all the links older than that to be omitted from the network. This is primarily useful when the effect of previous partnership in the network is examined, while choosing a new partner. For that, whenever a pair of scientists is chosen to work together, their link age is checked, and the more recently they have had worked together, the more probable it is for them to choose each other.

For the purpose of analysing the mutual collaborations in the Canadian biotechnology network, all the links in the two databases are analysed. For the links that occurred more than once then, the times between occurrences are measured and recorded. According to the databases, the average number of years between reconnection of links is 4.7 years. In the model, this fact is made effective while searching for a new partner, in the way that if the last collaboration between two scientists has been done in less than five years ago (because of the yearly basis of the model), the chance of choosing the partner increases by a fixed number.

In the procedure of choosing a new partner the following conditions are checked by the model: if the two scientists are from the same cluster, if the two scientists are from the same organization type, if the field of work for both scientists are the same, if at least one of the scientists is star, if scientists are Gatekeeper, if the distance between two scientists from two different clusters is less than a threshold, if the difference in the score number of two scientists is less than a threshold. For each condition, a fixed positive number is added to the chooser variable according to the level of advantage of the condition, i.e. if the scientists are from the same cluster they are more prone to work together, so a big positive number is added, while if the fields of them are the same a smaller number is added, since innovation is prone to occur in the collision of different fields of science as well. The overall flowchart of the simulation model is presented in Figure 2.

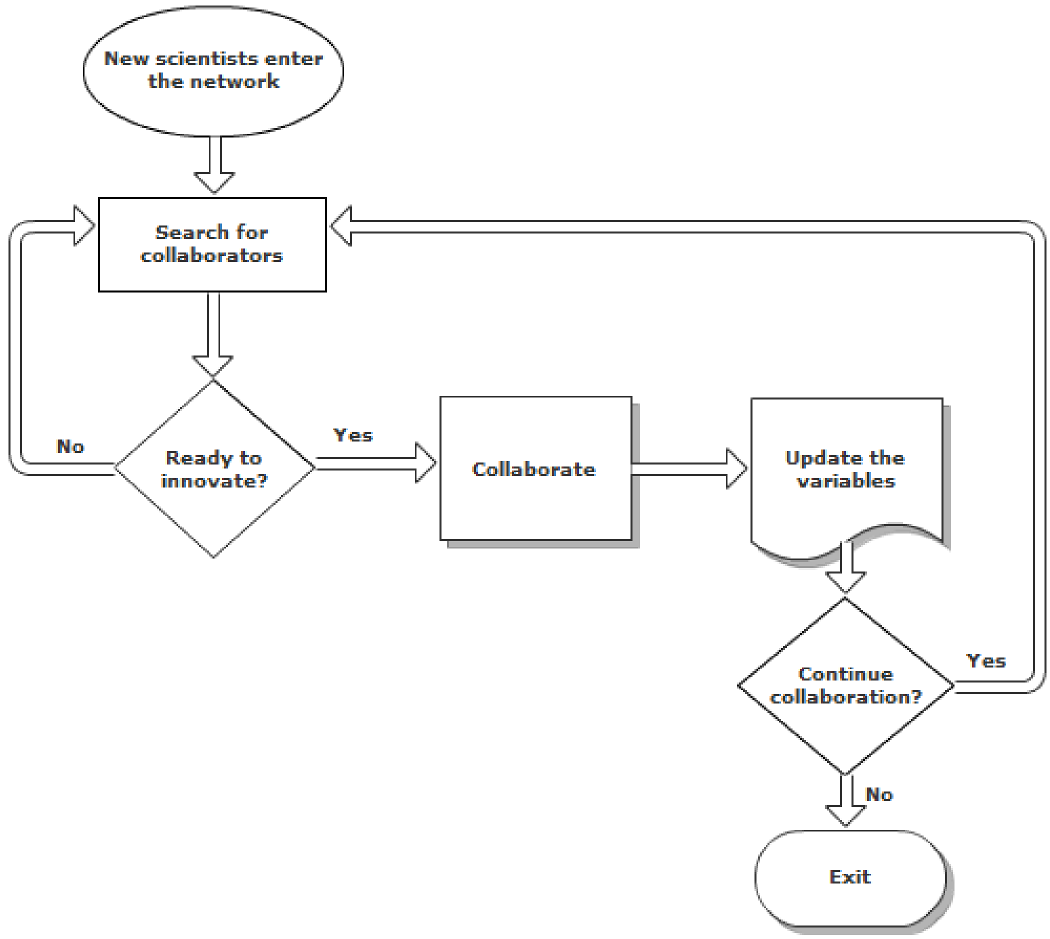

Figure 2: Flowchat of simulation model

More details on some attributes and variables, and where applicable input analysis procedures are provided below:

Geographical Cluster: In the database of articles extracted from SCOPUS, the information about the geographical locations of the affiliations of the authors is available. According to the database, there are a total of 160 countries. 58% of the scientists are Canadian, 19% American, and 23% from all other countries. Since the main objective of the present thesis is analyzing the behaviour of Canadian biotechnology inventors, the geographical location of Canadian scientists should be analyzed in detail, while the location information related to the foreign scientists could be simplified. For this purpose, the scientists are partitioned geographically. According to the database, fifteen geographical groups are defined, thirteen of which correspond to Canadian biotechnology clusters, one group includes all the scientists affiliated to the American institutions and one group puts together the scientists from all the other locations in the world. For the remainder of this thesis, these groups will be referred to as geographical clusters. The list of the geographical clusters and the percentage of scientists in each is represented in Table 2.

Table 2: Geographical Clusters

Geographical cluster	# of inventors	%
Toronto	22428	11
Montreal	21332	10
Vancouver	13154	6
Edmonton	7530	4
Ottawa	7466	4
Calgary	5447	3
Quebec	5936	3
KINGSTON	2833	1
SASKATOON	3772	2
WINNIPEG	5173	3
HALIFAX	3374	2
SHERBROOKE	2122	1
OTHERS CANADA	18180	9
UNITED STATES	38942	19
OTHER COUNTRIES	46163	23
Sum	**203852**	**100**

The collaborations between each pair of scientists are also analysed separately. The Table 3 shows some statistics related to the number of collaborations between geographical clusters. As it is shown in Table 3, most of the collaborations occur inside the cluster. This confirms the fact that scientists are more willing to collaborate with partners inside their clusters. The highest inside cluster collaboration is for Canada, which is predictable because the database is mainly established based on Canadian biotechnology collaborations. However, as the distance between two geographical clusters grows, the distribution of the number of collaborations becomes the same, i.e. 9.5% between Canada and United States, 8.9% Canada and other countries, and 5.3% other countries and United States.

Table 3: Distribution of collaborations between geographical clusters

Collaborations	Number	Percentage
Inside Cluster	1021861	71.20%
Inside Canada	694384	48.30%
US - Canada	136990	9.50%
Other - Canada	128349	8.90%
US - Other	75719	5.30%
Inside US	191565	13.30%
Inside Other	209181	14.60%

The simulation model is therefore designed in a way that scientists from the same geographical cluster are more likely to be chosen as partners. In the simulation model of the present thesis, the cluster of each scientist is an identity factor for him/ her in the network. Beside the identification, the main usage of the geographical location of scientists and their distances, which is basically dependant on their clusters, is in the decision making process for choosing a new partner. As it was also mentioned in the literature review (Cowan and Jonard 2004; Frenken et al. 2009; Schiffauerova and Beaudry 2008), the scientists from the same cluster are more willing to work together. Although it was not directly verified from the database, the literature also suggests that as the distance between two geographical clusters grows, the biotechnology collaborations are less likely to take place. The reason for this is that biotechnology innovation involves a highly tacit knowledge, whose diffusion is limited over long distances (Schiffauerova and Beaudry 2008). Therefore, the model applies a distance based factor, which causes the scientists from different clusters to be less willing to choose each other for collaboration. It is also suggested in the literature that as soon as the distance passes some minimum limit, it loses its importance, i.e. any scientists farther than 500 kilometres has the same probability to be selected (Schiffauerova and Beaudry 2009). Accordingly, the model is also designed in a way that for all the distances greater than the average distance between all pairs of clusters, the probability of selection becomes the same.

Since the gathered data encompass the period of about 55 years, there are instances where a scientist belongs to two different clusters at different points in time. There are two possible ways of dealing with this situation. The first option is to identify the scientists who have changed their location during this period, and record their cluster at each time of publication, and then make these variations effective in the model. However, since the number of scientists who have changed their location in the database is less than 3% of all, we can neglect this change. The second way of dealing with this situation is considering one scientist with two different clusters as two distinct scientists, one entering the network at a sooner time with cluster A as his/her identity, and leaving the model when the second one enters the network with cluster B as his/her identity. The second way is preferred in this thesis because it is more straightforward and leads to a clearer picture of the whole network. Moreover, calculating the network measurements (betweenness centralization, degree centralization, etc.) with a dynamic node would introduce additional complexity to the analysis. Since geographical cluster of a scientist affects the decision making procedure in the model by considering the distance as a drawback when choosing a partner, this hypothesis does not make any change in the results.

For assigning the cluster numbers in the thesis, first from the two databases a separate data sheet including the clusters and scientists' first and last names is developed. Then the redundant records are deleted, using SQL coding. This new data sheet excludes the duplicates of scientists or clusters. Then, the number of occurrences for each cluster is calculated, using SQL code. Finally, the resulted data is fed into ARENA input analyzer to get the best fit. The suggested distribution function for the cluster's numbers is as follows:

$$\text{Cluster} = 0.5 + \text{EXPO}\,(8.26)$$

Square Error = 0.0014

in which EXPO is the inverse probability function of exponential distribution. This equation is then used in the simulation model whenever a new scientist is added to the network, in order to assign cluster ID to him/ her by rounding the resulted value to the nearest integer number.

Organization type: The organization type of each scientist is also an identification factor for him/ her. Each scientist belongs to a certain organization type at each point of time.

The information regarding the organization type of each scientist is derived from his/ her affiliation in the database, which is included in the SCOPUS database and which was extracted to our database of articles. The affiliations were classified into five distinct organization types: firm, hospital, university, institution (i.e. governmental laboratories, research centers, etc.), and individual. The individual scientists belong to no organization at the time of the publication.

Like geographical cluster, the organization type for a scientist could also change during his/ her life in the network, meaning that one scientist may change his/ her organization in the same cluster or when moving from one cluster to another. The database is analysed to discover the probability distribution of scientists within various organization types. Most of the scientists (75%) remain within only one type of organization during their lives in the network. However, there are still 25% of inventors who have had publications under more than one organization type. In order to deal with this situation, any possible combination of organization type is considered as a new type by itself, i.e. for the scientists who have had both university and institution in their affiliations, a university/ institution type is defined.

In order to assign organization types in the simulation model to each new scientist, the percentage of each type of organization is calculated, extracting the data set of authors and organization types. The Table 4 shows the distribution of scientists in various organization type combinations. As it is shown, half of the scientists have publications solely under university affiliation, while one fourth of scientists publish either under university or institution affiliation.

Table 4: Organization types

Organization Type	Number of Scientists	Percentage
University	80725	50.30%
Institution	20966	13.10%
Institution / University	20239	12.60%
Hospital	11559	7.20%
Hospital / University	9592	6%
Hospital / Institution / University	6203	3.90%
Firm	5774	3.60%
Firm / University	2360	1.50%
Hospital / Institution	1340	0.80%
Firm / Institution	810	0.50%
Firm / Hospital / University	549	0.30%
Firm / Hospital	204	0.10%
Firm / Hospital / Institution	102	0.10%
Individual	50	0%
Sum	**160473**	**100%**

In order to analyse the collaborations between various organization types, all the mutual collaborations in the database of articles are studied to illustrate the preference of scientists in terms of the organization types of their partners. The results are shown in the Table 5. As it is illustrated in the Table 5, 73.46% of the entire collaborations occur between the partners of the same organization type. Besides, in more than 70 percent of the collaborations, at least one scientist is from University. The simulation model is designed in a way that in 73.46 percent of times, scientists are more willing to work with partners with the same organization type, while in the remaining 26.54 percent of times scientists choose to work with partners from other organization types according to the percentages provided in the Table 5.

Table 5: Organization Combination of Mutual Partners

Organization Combination	Percent of Collaborations
UNIVERSITY/ UNIVERSITY	48.9
INSTITUTION/ INSTITUTION	12.6
INSTITUTION/ UNIVERSITY	12.4
HOSPITAL/ HOSPITAL	9.0
HOSPITAL/ UNIVERSITY	6.9
FIRM/ FIRM	3.0
HOSPITAL/ INSTITUTION	2.7
FIRM/ UNIVERSITY	2.5
FIRM/ INSTITUTION	1.2
FIRM/ HOSPITAL	0.7
INDIVIDUAL/ UNIVERSITY	0.0
INDIVIDUAL/ INSTITUTION	0.0
FIRM/ INDIVIDUAL	0.0
HOSPITAL/ INDIVIDUAL	0.0
INDIVIDUAL/ INDIVIDUAL	0.0
Sum	**100**

However, since the total population of scientists from University type is very high, it cannot be concluded that scientists from the University have higher chance of being selected. Such comparison can be made when the populations of all kinds of organization types are the same.

Research field: Scientists carry out their research within various fields of specialization. In the real world, the scientific research interests of scientists are determined mainly by one's graduation field, and in the present thesis they are hypothesized to remain unchanged during the scientist's life in the network. One of the criteria in choosing the working partner for generating new scientific or innovative outcome is also the similarity of scientists' research specializations. The simulation model developed for this thesis therefore considers the similarity of research fields as a decisive factor for choosing the partner.

Biotechnology, as the main area of specialization for the network of the present thesis, is a vast field of study with numerous fields of specialization. In the two databases, scientists' fields of specialization could be tracked via their affiliation information. According to the affiliations and publications, four distinct general zones of specialization are distinguished in the database of articles, based on the keywords identified in their title or abstract. In order to assign specialization field to each node of the model, the co-authorships are analyzed respecting these four research fields, and corresponding percentages are extracted for each of them (first research area 59.61%; second research area 21.68%; third research area 14.65%; and fourth research area 4.06%). For the present model, it is hypothesized that all four fields have equal importance as to be chosen for co-authorships. This means that as soon as a scientist finds someone with a different specialization field from his/ her field, a positive value is added to the decision variable, making it more probable for the two to work together. However, the simulation model is built flexible and the total number of specializations can be set as required in further utilizations of the model. Since no trend is detected in the two databases for the research fields, this attribute is randomly assigned to the scientists in the model, based on the calculated probability of occurrence for each field.

Chooser threshold: In order to make each scientist' decision making process unique, an attribute is defined for each scientist separately as his/ her *choose threshold.* This number is a uniform random number in the model, which remains constant during the life of a scientist in the network, assuming that preferences of any scientist for choosing a partner do not change during his/ her life.

In the model, some hypothesised decision making factors are used in order to imitate the behaviour of different scientists in the network. In the real world, scientists as individual human beings may have very personal motivations and reasons for choosing to work with others. Detecting all the incentives and recording them requires questioning each individual about their specific criterions separately. Developing such a detailed list of all the incentives requires direct questionnaire for each individual scientist, which is not feasible for the purposes of the present thesis, since only the network related characteristics of the scientists are important. In the present thesis, the decision making process has been based on the available data. According to the databases, the decision criteria of scientists about their partner's network-based characteristics can be traced by analysing the mutual collaborations in the network. For example, as it was shown in the "geographical cluster" section, almost 71% of the scientists prefer to have partners inside their own clusters.

The decisive factors that were traced from the databases are: geographical cluster, organization type, history of previous partnerships, research field, Gatekeeper status, star status, and scientific quality of each scientist. Therefore, the model is designed in a way that all these factors are considered when choosing a new partner. For this purpose, two temporary variables called *preference indicator* are introduced in the model whenever the model is considering two scientists for a possible partnership. The model checks each of these decisive factors between the two scientists, and updates the *preference indicator* variables with a new value as soon as each criterion is met. For example, if the two scientists are from the same cluster, then a positive value is added to both *preference indicator*s, and when one of the scientists is star, then a positive value is added to the other scientist's *preference indicator*, representing his/her desire to work with a star.

In the process of new partners selection, first each two scientists (potential partners) are examined in terms of all the decisive factors, and the references of each towards his/her possible partner are summarized by his/her *preference indicator* variable. Then the *preference indicators* of each of the two scientists are compared to the *chooser threshold* of each scientist, and as soon as the *chooser threshold* of both scientists is met (i.e. for each of the two scientists, his/her *preference indicator* is greater than his/her chooser threshold), they are chosen to start a co-publication together.

Entrance year: As it can be seen in the databases, scientists enter the network in some period of time and produce innovation (i.e. scientific articles or patents), and then leave the system either temporarily or permanently. Since it is not expected from a scientist to be active forever, the entrance year for each scientist is recorded. This allows him/her to exit the network after a certain period of time, even if he/she is very productive. In other words, the entrance year variable keeps track of the age of each scientist in the network, and let him/her exit the system after some average number of active years, just the same as in the real world. In order to calculate the average age of a scientist, the maximum and minimum years of activity of each scientist is extracted from the two databases. According to the results, a maximum number of active years are 37, whereas a minimum is 1 year. The data is fed into ARENA input analyzer to develop a mathematical formula for the age of scientists, and the results show that age of a scientist in the network behaves exponentially, with a mean of 5.66. This formula is then applied to the model to produce random ages for the scientists.

For calculating the number of scientists who should enter the network at the beginning of each year, the two databases of patents and publications are used. First, the data regarding the name of each scientist and year of publication/patent grant are extracted from the two databases and merged together. Then, using SQL code, only the first year that the name of each scientist appears in the data is recorded separately. This year actually indicates the entrance year for the scientist in the Canadian biotechnology network. After that, in order to develop a mathematical distribution function for the trend of entrance, the numbers of duplicate years are recorded. The result is then fed into input analyzer of ARENA software in order to get the best fitting distribution. The result is as follows:

$$\text{Entrance year} \; = \; 1.95\text{e} + 003 \; + \; 55 \; * \; \text{BETA}\,(0.958, 0.978)$$

Square Error = 0.000499

in which BETA is the inverse probability function of Beta distribution. This function is then fed into the model, in order to generate entrance years for scientists.

Number of publications/patents: As soon as a new scientist is created in the network, he/she should have some primary characteristics assigned in order to survive in the network of the simulation model and get engaged in the decision making process for choosing new partners. It is therefore hypothesised that some scientists enter the network from other already existing networks, which enables them to have had some initial number of publications and patents before entering our network. It should be noted that this primary number is only used for modeling purposes and does not affect the whole procedure and/or simulation results.

There is no exact evidence in the two databases as to the previous publications of authors entering the Canadian biotechnology network, so the model is designed in a way that this primary number of publications is flexible and for further analyses could be set as required. This initial setting does not affect the overall simulation results, but shortens the warm-up period of the model. Since the model is designed as finite horizon, it is necessary to have some decision making criteria enabled at the beginning, so that a shorter warm up period towards stable results is achieved. In the model of the present thesis, the two numbers (the initial number of the articles and of the patents) are set as a uniform random number between 1 and 10 for each scientist. However, the initial number of publications does not affect the final analysis for this thesis, because only the rate of productivity of the network is of importance, and this is detectable trough a cumulative trend of number of papers and patents in the network.

Although the primary number of publications and patents does not affect the innovative productivity of the network, it can be used as a decisive variable while choosing to work with other scientists in the network.

After any co-authorship or co-inventorship event the total number of publications/patents for each scientist in the group is updated. First the model determines the type of the event to be either an article publication or a patent registration. According to the two databases, 3.53% of all events are patent co-inventorships, and the remaining are article co-authorships. Then, if it is an article, the number of publications for all the co-authors will increase by one. If it is a patent, then the patent quality of all the co-inventors will be updated according to the patent quality distribution in the USPTO database. The patent quality is based on the number of claims[11] for each patent in the database. The probability distribution function for the patent quality is built in ARENA input analyzer as follows:

Expression: -0.001+GAMMA (10, 1.59)

Square Error: 0.001

in which GAMMA is the inverse probability function of Gamma distribution.

Primary links: Similarly, it is assumed that every scientist has a random number of links initially. Again, this assumption will help the model to pass its warm up period faster. The input data for this model is based on a finite horizon, meaning that the data has a starting and ending state. According to the database, each scientist has some number of links as soon as he/she enters the network, since the links are defined based on the co-publications. In the database, the number of links inside the cluster per scientist is between 1 to 5 links, and outside the cluster it is 0 to 2 links. In other words, a scientist

[11] Patent claims are numbered expressions used to describe the technical terms of invention and demonstrate the protection extent presented by a patent. A high number of patent claims indicates that an innovation is broader and more profitable, potentially (Schiffauerova 2009).

may have 1 to 5 links inside its cluster and 0 to 2 links outside its cluster as soon as he/she enters the network.

The average number of initial links for each scientist is derived from the database. For that, the first 10 years of data is analysed and the average number of links per scientist is calculated separately. Then the probability for each number of links per scientist inside and outside of the cluster is calculated separately. Table 6 shows the percentage results for the first ten years.

Table 6: Number of links per scientist inside and outside of the cluster (%)

Number of Links	0	1	2	3	4	5	sum
Inside Cluster	6%	18%	31%	12%	25%	8%	100%
Outside Cluster	44%	37%	18%	1%	0%	0%	100%
Sum	50%	55%	49%	13%	25%	8%	-------

As mentioned before, the initial number of links assigned to each scientist helps in reducing the simulation running period by omitting the warm up period. Since the simulation model is agent based, the initial picture of the system should be comparable with any other picture taken in various time slots during the replication. In addition to reducing the warm up period, this realistic initial status of the model ensures that the whole output data of each replication is appropriately the same as the real behaviour of the database.

Journal impact factors: Since in the world of publications it is not only the quantity which counts, the quality of the scientific output should be taken into consideration as well. The quality of the journal articles produced by the scientists should therefore be evaluated. The article forward citations are so far generally recognized as the most appropriate paper quality indicator, but unfortunately the information on the article citations is not available in our data. Instead, it is decided to include the journal impact factors corresponding to the journal in which each article is published in the model. Journal impact factor is admittedly a very noisy indicator of article quality, but it does give a certain indication on the scientific value of the work. The impact factors of the journals for the year 2008 are extracted from ISI Web of Knowledge[12]. The corresponding impact factor for each journal in which the article has been published is then assigned separately to each article. Using ARENA input analyser, the following expression is provided for the journal impact factor distribution in the data.

Expression: -0.001 + LOGNORMAL (2.3, 3.73)

Square Error: 0.0015

in which LOGNORMAL is the inverse probability function of Lognormal distribution function. Using the above mentioned formula, as soon as a new article is produced, some random number as journal impact factor is added to the corresponding variable for all the co-authoring scientists.

The histogram presented in Figure 3 shows the frequency of journal impact factors for all the articles in the database. As the figure show, most of the articles (more than 50%) are published in journals with impact factor less than 2.5.

---

[12] An academic search service for citation indexing, provided by Thomson Reuters (www.isiwebofknowledge.com)

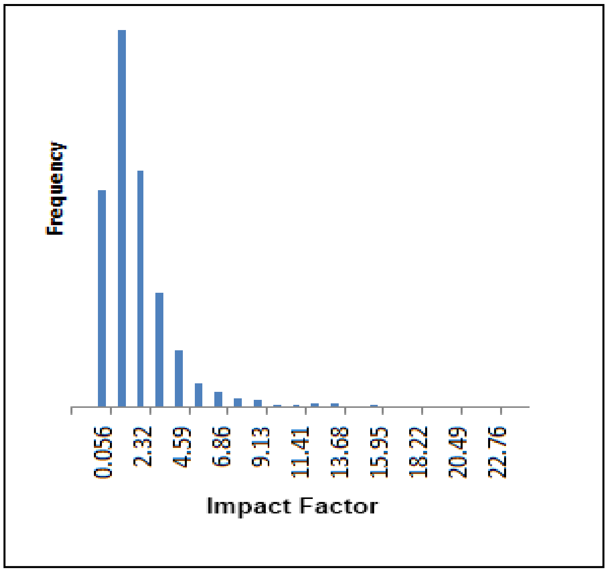

Figure 3: Frequency of Journal impact factors for articles

Idle/busy status: There are two general statuses defined for each scientist in the network: busy; and idle. Whenever the status of a scientist working on a co-publication is set as busy, he/she cannot be selected for any other co-publication and is neglected in the chooser procedure of scientists during this period. However, in the real world, it is possible for one scientist to work on more than one publication or innovation at the same time. This means that even while working on a project and being busy, in the real world a scientist still could be contacted for working on other projects. This phenomenon is not detectable through the databases, because in the both databases only the years of article publication in a journal or innovation registration at the patent's office are recorded. There are many scientists in the databases who have published more than one journal paper in one year. Even though it is not clear from the database whether the periods of work on these publications have had any overlap. It is expected that in order to produce a research article at least several months of research is required and thus the simultaneous work on several collaboration projects is likely. Nevertheless, in the simulation model of this thesis, it is assumed that each scientist can only work on one single project at most, at any point in time.

This assumption affects the possibilities for choosing a new partner, as the scientists who are already working on a project are opt out of selection. The fact that each scientist can work on one single project at a time also affects the number of links a scientist has in the network. Although scientists in the real world usually work on more projects at the same time and thus should have a higher number of simultaneous links, the assumption reduces this number in the simulation model. However, this situation is solved by the defined age for the links in the model. As it was stated before, as soon as a link is connected between two scientists in the network, this link remains active in the model for five years. Now, if a scientist in the real world works on more than one innovation at a time and has more number of links connected to him/ her at the same time, the situation is solved by this link age.

The database of the simulation model results is produced in MS Excel for each run, which has a limited capacity for the size of the matrix of nodes in the network. Besides, as the number of links per scientist increases, the run time of model increases as well. In order to speed up the model and overcome the limits of the software, and in order to make the scientists have realistic number of publications in the network, a monthly time frame is considered in the model. As it was stated before, the database records only bear the year of publication, and it does not indicate the publication duration. Therefore, if the publication duration in the model is set to be less than a year, then each scientist has the opportunity of producing several publications in a year, which could imitate simultaneous co-publications for each scientist.

In order to have a more realistic picture of the network for further analyses, the realistic number of links per scientist is required. The solution proposed for this situation in the present thesis is to develop a prolonged picture of the system for output analysis purposes. This means that instead of having a picture of the nodes and links in the network at each single point of time, the model would give all the information about all the links and nodes over a longer period of time. This picture is more realistic when this period of time is set as a year, since the database on which the model is developed proceeds yearly.

In order to extract the number of busy scientists for each year, first the two databases are merged together, and the query of publications, years and scientists is developed. Since redundant entries are possible in the merged query, a SQL code is developed to produce the distinct query of scientist, year and publication, in which each scientist is only once related to each article he has published during a year. Then, duplicates of scientists for each year are extracted from the resulted datasheet, using wizard of MS Access. The Figure 4 shows the growth rate of number of busy scientists during the period under study.

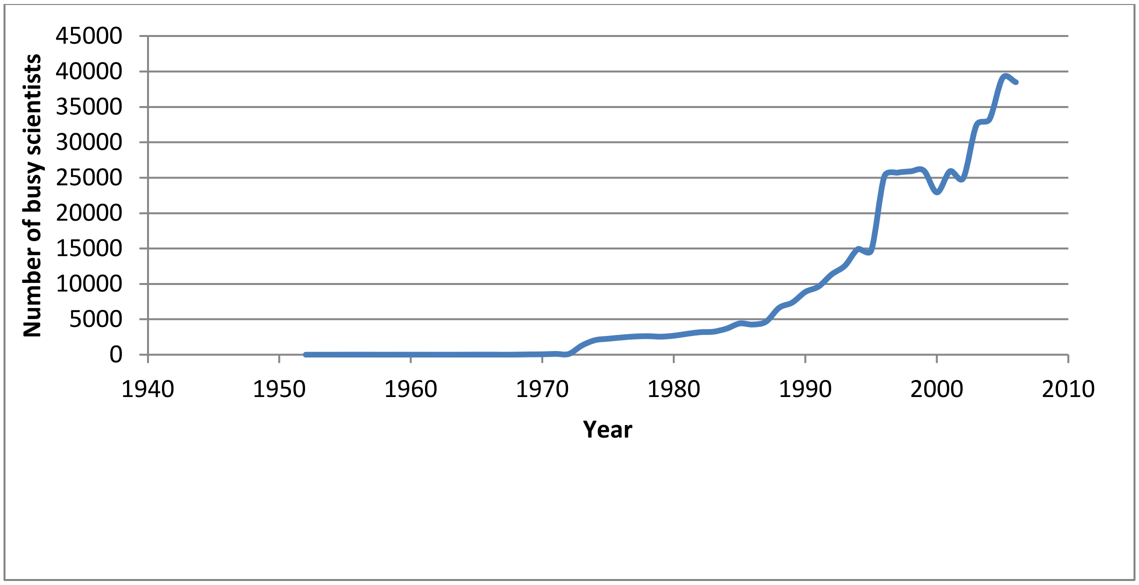

Figure 4: Number of busy scientists per year

As it is seen in the figure, the growth rate of scientists in the network is very fast. The figure shows that the biotechnology as a new and separate field of study was initially introduced in 1970s and the beginning of 1980s, and that there were not many patents and publications before that. However, as the figure also suggests, the growth rate of number of scientists in Canadian biotechnology network is high, and the number of scientists in this field is still growing at the end of 2005. This implies that in the simulation model of the present thesis there should also be a growth in the number of scientists working in this field as the time passes.

Feeding the data in ARENA input analyzer the best match is as follows:

$$\text{Number of busy scientists} = -0.001 + \text{EXPO}(19.6)$$

Square Error: 0.015135

In which EXPO is the inverse probability function of exponential distribution. The square error is relatively high, since the data is very scattered in the first years. This probability function is then fed into the simulation model whenever a scientist should decide to start a new publication, i.e. going from idle status to busy. The model checks to see if the number of current busy scientists is less than the number produced by this function, to add a new busy scientist to the model.

Publication duration: The main advantage of simulation modeling, as stated before, is the ability to analyse systems over time. As soon as one wants to build a simulation model, the exact timings for events in the system should be recorded to be made effective in the model. In the present thesis, the most important event in the network is the process of co-publications of the articles and co-inventorships of the patents between the scientists. In order to have an exact formula for the publication duration however, the exact information about the beginning and end of each publication is required. In the two available databases, the only time value available is the year of publication or patent grant. Therefore, other time elements should be estimated based on other observations of the data.

For calculating publication duration, the number of publications per scientist per year can be considered. As long as a scientist produces only one publication during a year, it can be assumed that the publication duration is one year. However, for imitating the publication process of scientists who have worked on more than one publication at a time, the query of scientists with more than one publication during one year is taken from the databases. Then, the probabilities for each number of publications per year are calculated separately. Based on the possible number of publications, various possible publication durations are then defined in the model. This means that for example, for two publications per year, the probability is assigned to the six month publication duration, or for three publications per year the corresponding probability is assigned to four month publication duration, and so on.

Therefore, as soon as a group of scientists is selected to start a new research contribution in the network, specific publication duration is set for them, during which their status is set as busy, while before and after that their status is idle. This way, for example, if a scientist in the real world has had three simultaneous publications at the same time, he/she is represented in the model as someone who has three consequent publications during one year. This regulation also helps the computer model to run faster, as the number of choices that should be considered each time for starting a new publication is confined to the idle scientists only. The results for the behaviour of both the simulation model and the real world (represented in the present thesis by the database) would be the same, since at last both of the results are presented on a yearly basis. However, if some more detailed data is provided in the future, the simulation model of the present thesis is defined flexibly and it could take new settings for having more than one contribution at a time.

For calculating the probabilities for each busy period, first the query of scientists, publications/patents and years is extracted from the databases. The total number of publications in comparison with patents is very high, i.e. 100750 articles versus 2932 patents. Therefore, the probable difference between the duration of patenting and publishing can be neglected, and both data sets can be merged together for duration calculation purposes.

Since the two databases of patents and publications are combined together for this purpose, some duplicate records may occur. To overcome this problem the duplicates are removed by using SQL codes. Then, using MS Access query wizard, the duplicate years for each scientist are calculated. The Figure 5 illustrates the percentages of publication duration and patenting duration taken from the two databases.

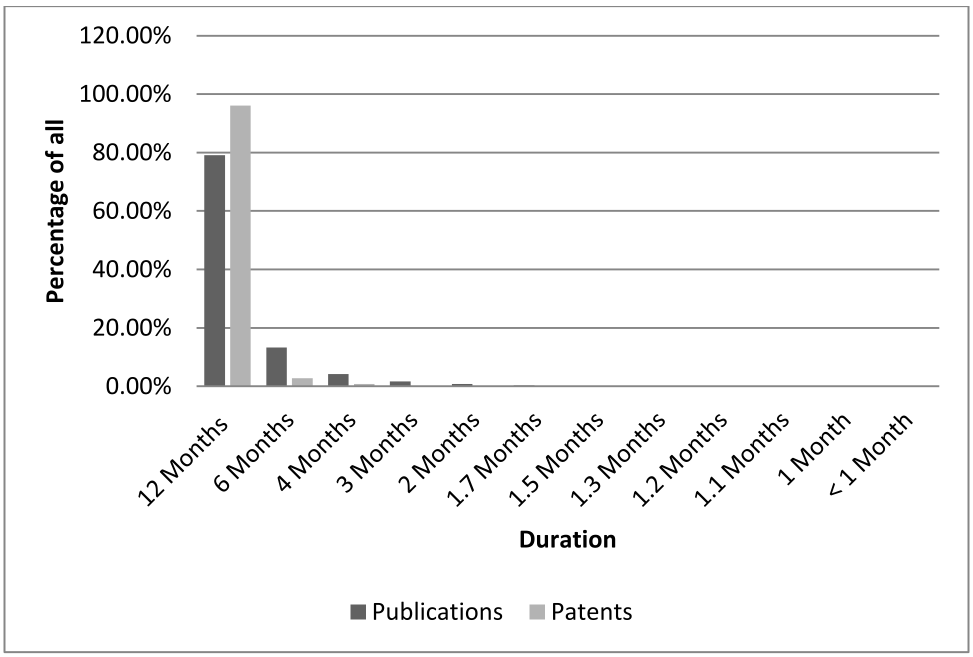

Figure 5: Publication/ Patent Duration

As it is illustrated in the figure, almost 80% of the scientists have had only one publication per year, and for more than 95% of times, number of patents per year is also one. As the number of publications per year increases, the total number of incidents in the data falls significantly. Therefore, the simulation model is designed in a way that about 80% of the times the publication duration is set as one year, which means that the status of each scientist involved in the publication is set as busy for one year. Other possibilities of publication durations and their corresponding probabilities are also introduced to the model based on the developed equations.

Number of co-authors: After the initial attributes and variables are set in the model, the simulation replication starts by searching for scientists with idle status. For each idle scientist in the network first the model decides whether he/she should start a new publication or not, according to the distribution of number of busy scientists per each year. The decision process for this purpose is randomized exponentially. If a scientist is chosen to start working on a new project, first a decision variable is needed to determine the number of co-workers.

In order to calculate the number of co-authors working on the same project, the database of articles is used. In this database, first a query of all the publications and all the scientists working on each publication is extracted. Then in the new data sheet, the number of duplicates for each paper is developed separately, using SQL. The distribution of the number of scientists per publication is depicted in the Figure 6. As the figure shows, about 80 percent of times, less than seven people are working on the same publication.

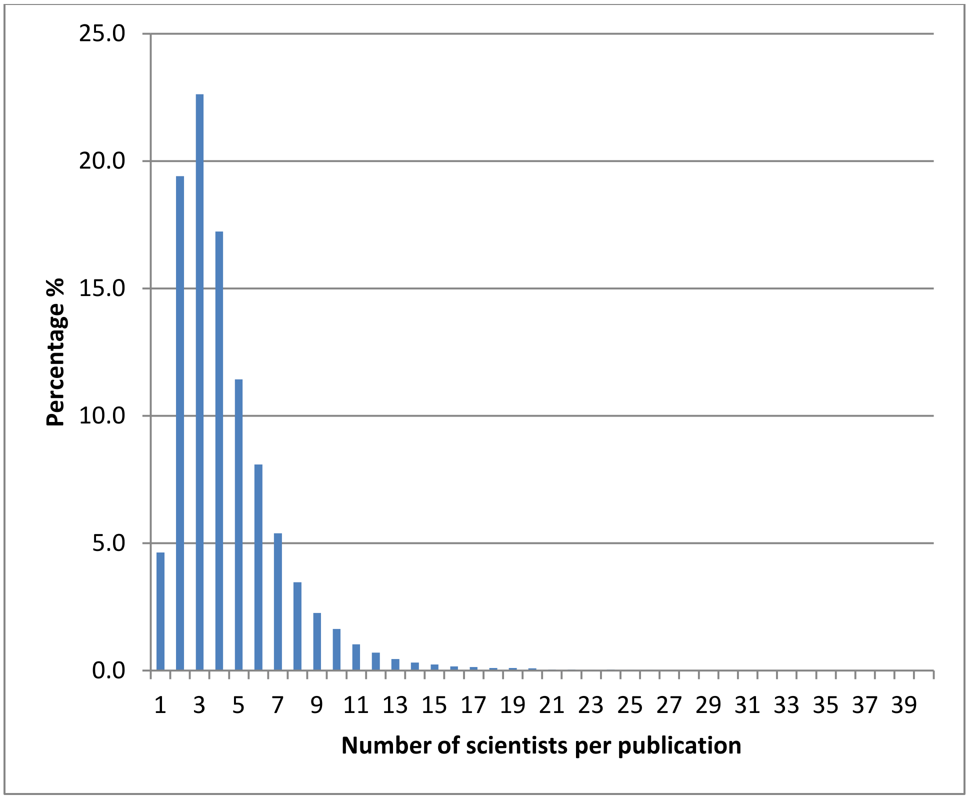

Figure 6: Distribution of number of scientists per publication

The simulation model is designed in a way that as soon as a scientist is chosen to start a new publication, a random number is generated based on the distribution above to decide about the number of co-authors needed for the new publication.

Star Status: According to Queenton and Niosi (2003), Canadian biotechnology clusters are strongly related to high-class academic research and especially to the star scientists working in the network. As it is mentioned in the literature review, there is no universal definition for star scientists that contains numerical specifications regarding their productivities. In this thesis, a highly prolific scientist is considered to be a star scientist, and it is assumed that he/she is more prone to be chosen by other scientists as a partner because of his/her scientific background and research success. Clearly, the higher the number of publications and/or patents a scientist has produced, the more he/she has been chosen as a partner and engaged in co-publications or co-inventorships.

There is no exact evidence in the databases which shows that star scientists are chosen because of their higher productivity. Decisive effect of star status of a scientist is as undetectable as other personal decision factors in the network. However, the model is designed in a way that as soon as a scientist meets some lowest specifications (minimum of 5 publications, 2 patents, and 10 journal impact factors, which forms about 5% of the scientists in the real world data), he/she is less prone to leave the network. In order to consider both qualitative and quantitative productivity of the star scientists, there are three criterions defined in the model for a scientist to be a star: minimum number of publications, minimum patent quality, and minimum article quality. As it is stated before, the model is flexible and the criterions are manageable for various runs of the model.

The following figures show the distribution of number of publications and patents for inventors in the Canadian biotechnology network (Figure 7). As it is depicted in the two figures most of the scientists produce less than 2 patents or 5 articles.

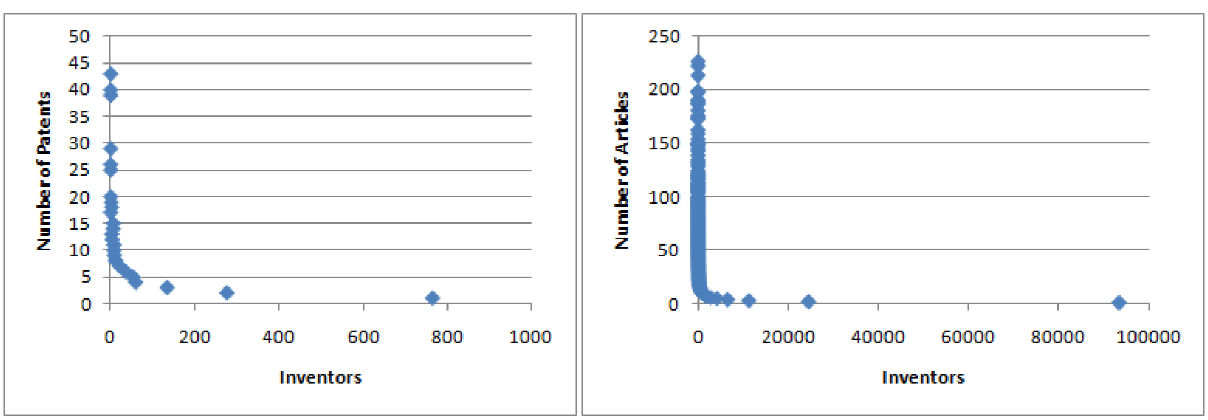

Figure 7: Number of Patents and Articles per Scientists

In order to calculate the number of stars in the databases, the list of the scientists with more than 25 publications and more than 10 patents are extracted from two separate databases. Then, the mutual scientists are detected in the two new sets of scientists. Here, among 49 scientists with more than 10 patents and 1955 scientists with more than 25 publications, only 33 stars are detected. Clearly, as the criterions are changed to 30 publications and 15 patents, the number of stars reduces to 11 stars in the network. The average journal impact factors in both cases are set more than 10.

In the model, the default values for minimum number of articles and patents for a star are set as 5 and 2 respectively. Besides, the minimum cumulative impact factor is set as 10. These specifications correspond to 4.7% of all the scientists in the databases. All these numbers are changeable for further analyses. At each point of time, the model checks the star status of a scientist, i.e. whether or not the scientist is a star in the network. As soon as a scientist meets the minimum specifications in the model, his/her star status is set as true. As it was mentioned before, the star status of a scientist plays an important role in the procedure of choosing a new partner, because in the model scientists are more willing to co-author with stars.

Gatekeeper status: As it was defined in the literature review, the term Gatekeeper in the present thesis is applied to any scientist with at least five links inside and two links outside its geographical cluster. These criteria correspond to almost 14.2% of all the scientists in the two databases. The same criteria is applied to the simulation model, in which any scientist with at least five active links inside the cluster and two active links outside the cluster is considered as a Gatekeeper. An active link is a link which has been in use at least once in less than 5 years ago.

In order to analyse the distribution of Gatekeepers, the database of the co-publications is analysed. According to the database, 14.2% of all the scientists have at least once been a Gatekeeper, i.e. they have had at least one publication with five inside and two outside collaborators. 73% of these Gatekeepers are Canadian, 14% American, and 13% from other zones. The Table 7 shows the distribution of number of Gatekeepers with their frequency in the database. As it is shown in the Table, 69% of all the Gatekeepers have only been Gatekeeper once, while the remaining 31% have played the Gatekeeper role more than once during their activities in the network.

Frequency	# of scientists	%
1	34156	68.5093068
2	8110	16.26684852
3	3112	6.241976893
4	1586	3.181161746
5	813	1.630696406
6	550	1.10317715
7	378	0.758183569
8	243	0.487403723
>8	908	1.821245186
Sum	**49856**	**100**

Table 7: Frequency of Gatekeepering

As it is stated before, the simulation model is designed in a way that the distribution of Gatekeepers in the model is almost the same as that of the database.

The code consists of four classes, Manager, Scientists, Cluster, and write excel. The manager is the main class of the codes. The variables and parameters in the program are named in relation with their functionality in the model, in order to make the Java code more understandable. Besides, some comments are made through the code wherever necessary to describe the functionality of each part of the code. Since the model is created for the purposes of this thesis, no user interface is defined for it, and all the variables and parameters are set directly inside the Java code.

In the next part, the verification and validation processes of the model are explained.

4.5 Model Verification and Validation

Model verification and validation, as two important phases of modeling, are basically integral sections of model development procedure. In the present thesis, as the whole simulation model is developed by coding in Java language, the verification and validation procedures are almost always done during the model creation procedure. In the following sections, the applicable verification and validation methods for the model and their results are described.

Verification of agent-based simulation models, just the same as other modeling methodologies, mainly involves debugging the model to ensure it works correctly. As for validation, which ensures the right model is built, Klügl (2008) proposes four principal methods for agent based methods, which are face validation, sensitivity analysis, calibration and statistical validation. However, Fortino, Garro, and Russo (2005) propose that discrete event simulation methods of validation can also be performed for agent based models.

The simulation model created for the present thesis is primarily defined as a finite horizon model, since the duration for which data is available is finite, and further evolutions after the last year in the two databases of the system are unknown. However, the simulation model is created in a way that it can be run over infinite horizons as well if required, as the Java code of the model is open source and may accept any positive number for the total number of years that the model is run.

Among all the possible terminating conditions for the model, the total number of years is selected as the termination point, since this is the only value that remains unchanged under all possible scenarios. Since the available data is gathered over approximately fifty years, any iteration of the model is also set to be fifty years.

The output analysis for the simulation results of the present thesis is based on the method of independent replications, given that all the variable generators utilized in the model are random. The output analysis for the model will be done according to the across replication method, meaning that for each set of results required for different research questions of the thesis, the simulation model is run a couple of times, and the average of all the results are going to be applied in interpretation procedures. The number of replications required for each set of results is calculated based on the t-distribution iterative method. The confidence interval for all output analyses is set as 95 percent in the software.

Verification: Verification of a software based simulation model actually means ensuring that the whole model is constructed correctly. Since the programming language of the simulation model is Java, which does a lot of debugging by itself, as soon as the model runs without any error notifications, one can be sure that the software phase of the model is correctly built. Besides, the model is verified by two programming experts for correctness of codes and consistency of the results.

> ***Static Verification:*** As for the static testing for the function of all subsections and modules of the model, a walk-through variable was defined so that it reported its situation as soon as it entered any module of the model. Using this walk-through variable, the correctness of the structural properties of the model was examined and confirmed consequently.
>
> ***Dynamic Verification:*** As for the dynamic testing of the model, the time flow mechanism of the model was verified by tracing an entity during its life in the model. The software based program showed no sign of inconsistency. The time flow was also checked to work monthly, as defined for the model, and all the sub modules were also verified to work on a monthly basis.

Validation: Validation phase of modeling is in fact assuring that the model which is built exactly imitates the real world phenomena. In the present thesis, the correctness of the model basically depends on the closeness of its behaviour to the evolution of the two databases in time. However, the two databases do not contain separable data to be used for model development and validation independently, so the same data is used for both model generation and validation. Besides, since the available data is heterogeneous in time, meaning that there is a significant change of behaviour in the state of variables such as number of scientists or publications over time, it could not have been divided into two time-based sections as well. For validating the model, the following validation methods are considered as applicable and applied to the model as explained hereby.

- ***Face Validation:*** The very basic validation method for a software based model is the usage of animation elements in order to compare the external appearance of the model to its instance in real world. As for the innovation network under study in the present thesis, the main elements that could be detected in animation are the scientists (nodes) and the links between each pair of scientists that lead to the creation of new publications. The animation interface for the model is mainly developed using various steady state pictures of the model, developed in PAJEK[13] network analysis software. The Figure 8 shows an example of the picture developed in PAJEK out of the simulation model of the thesis.

Since the total number of nodes and links in the real world network is too high to be pictured, for getting a better demonstration of the results for validation purposes, the

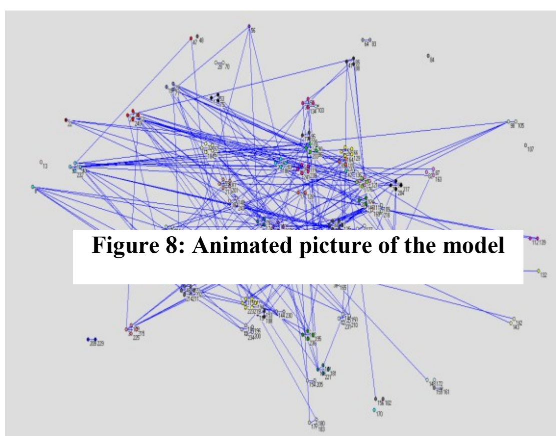

Figure 8: Animated picture of the model

simulation is run with a proportionally smaller population. For this purpose, all the relative settings have been set as one tenth of their correct values.

---

[13] PAJEK is a program for analysis and visualization of large networks, developed by Vladimir Batagelj and Andrej Mrvar.

As the figure suggests, the overall appearance of the model is acceptable, since each scientist has its own cluster, and the links are also generated properly for each pair of scientists working together.

- ***Degenerate tests:*** in order to validate the model, some degenerating tests are performed in the model.

– The score threshold in the model is set to a number bigger than the summation of all choosing factors in the model. It is expected that by applying this setting, no scientist can be ever found for partnership and therefore the number of links, and consequently scientists in the network decline to zero after a couple of years. Under this condition, as it is expected, the population of scientists in the model declines very fast. The Figure 9 depicts the situation over fifty years.

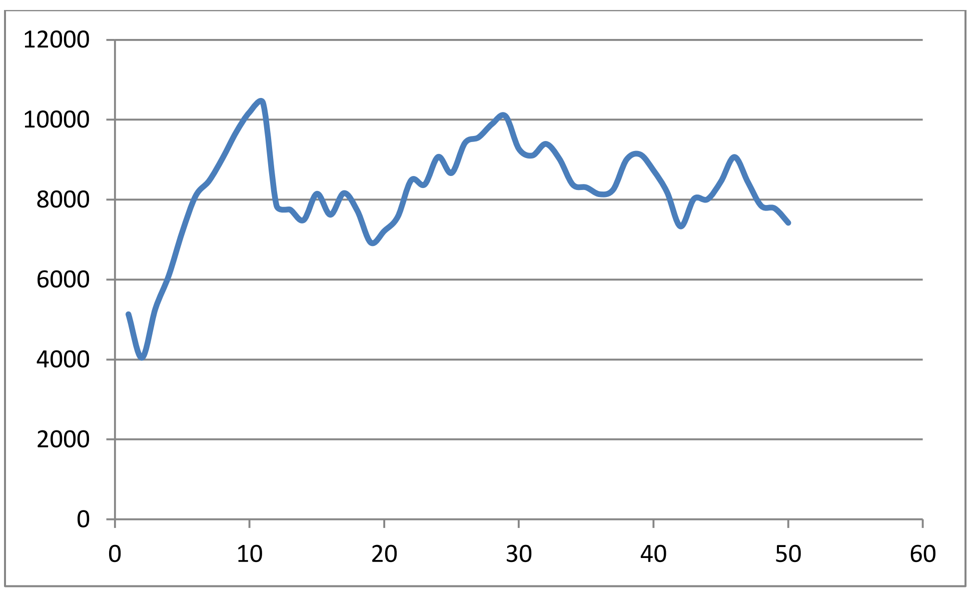

Figure 9: Degenerate test. No partner selection possible in the network.

– When the maximum possible age of a scientist in the network is set as one instead of twenty years, it is expected that the model stops working after a couple of years. As expected, the population drops to a number near zero in the first year, and the model does not proceed after the third year since there is no more any scientist in the network.

– Setting the maximum age of a scientist to a number greater than the whole duration of a replication, it is expected that the population of scientists grows positively in the model, since no scientist leaves the network during this period. The Figure 10 compares the number of scientists in the two situations. For comparison, the figure also illustrates that when the maximum age of a scientist is set as twenty years, the population of the model shows a steady behaviour.

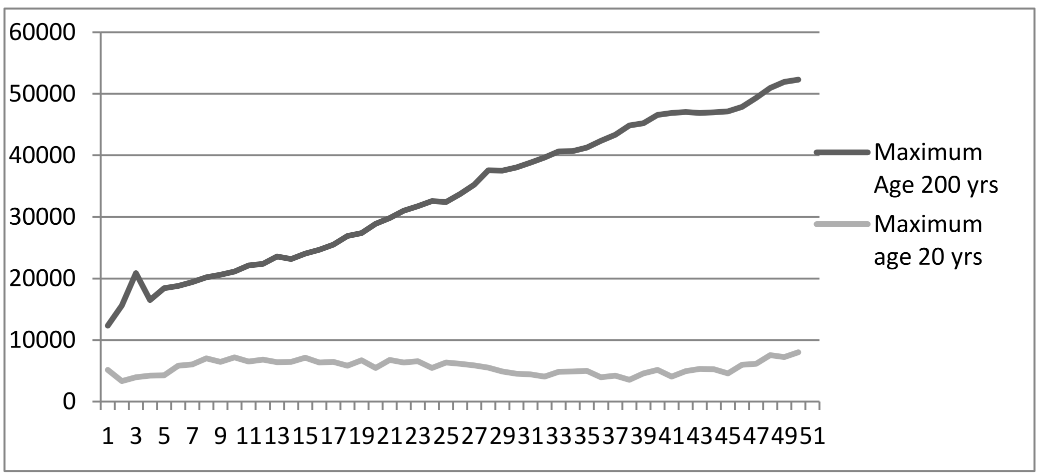

Figure 10: Degenerate test. Age of scientists

– It is also expected that the number of scientists declines fast if the probability of death in the network is set as a big number. The results show that with the high probability of fifty percent for unexpected omission, the population of scientists decreases very fast.

- ***Internal validity (Statistical Validation):*** For conducting the internal validity tests on the model, the same setting is run ten times in order to check the consistency of the results. Since different sets of random variables are used for each iteration, the produced data are supposed to be slightly different. However, the trends in the outcome data for iterations should resemble each other. The following figures show the trend of output data over one hundred years of run for five independent replications. The fact that the trends are the same confirms that the model is working heterogeneously.

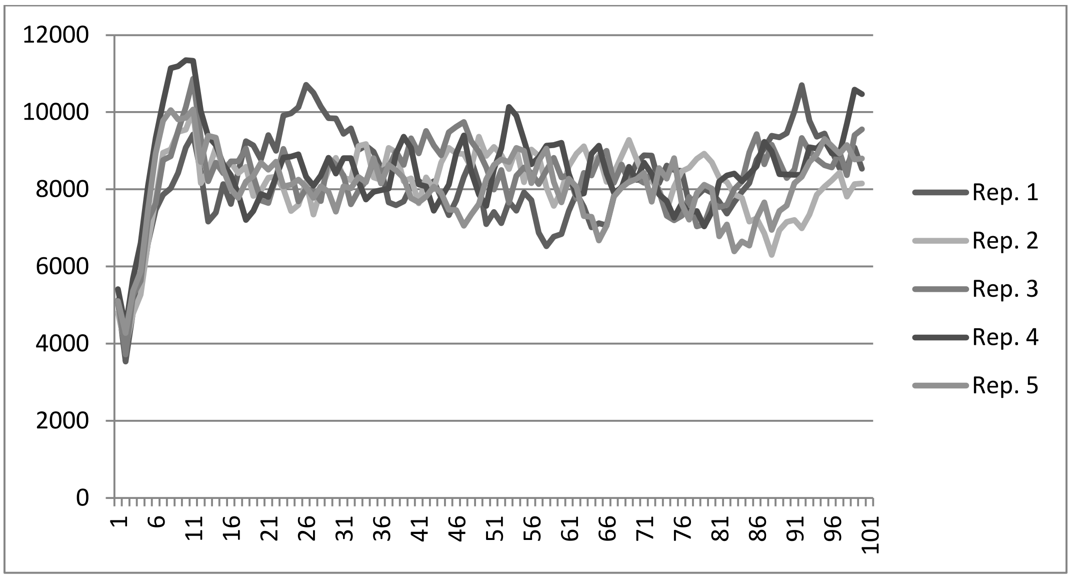

Figure 11: Internal validity. Data trend for number of active links in the network

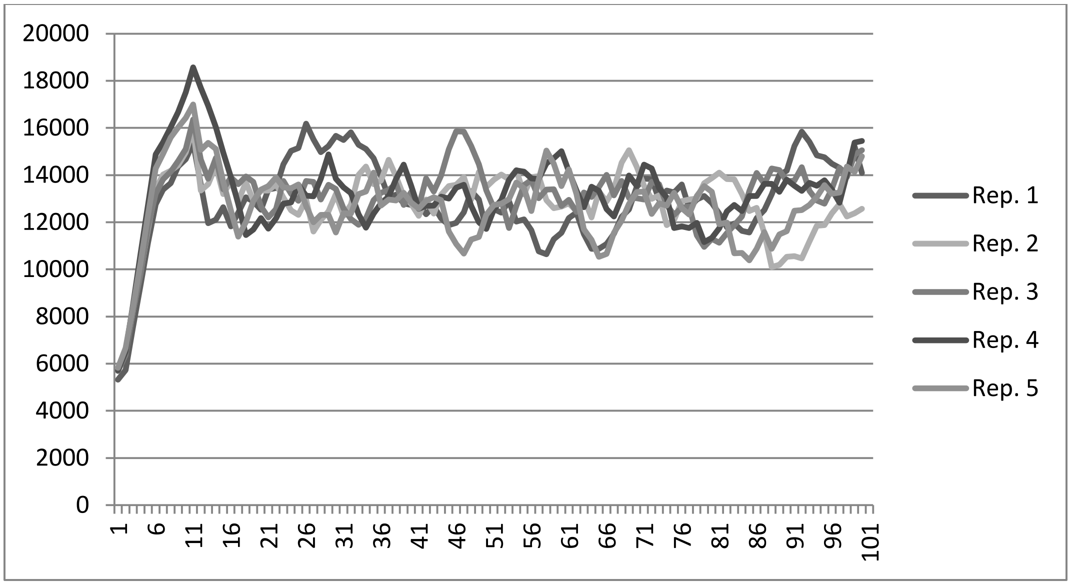

Figure 12: Internal validity. Data trend for number of publications

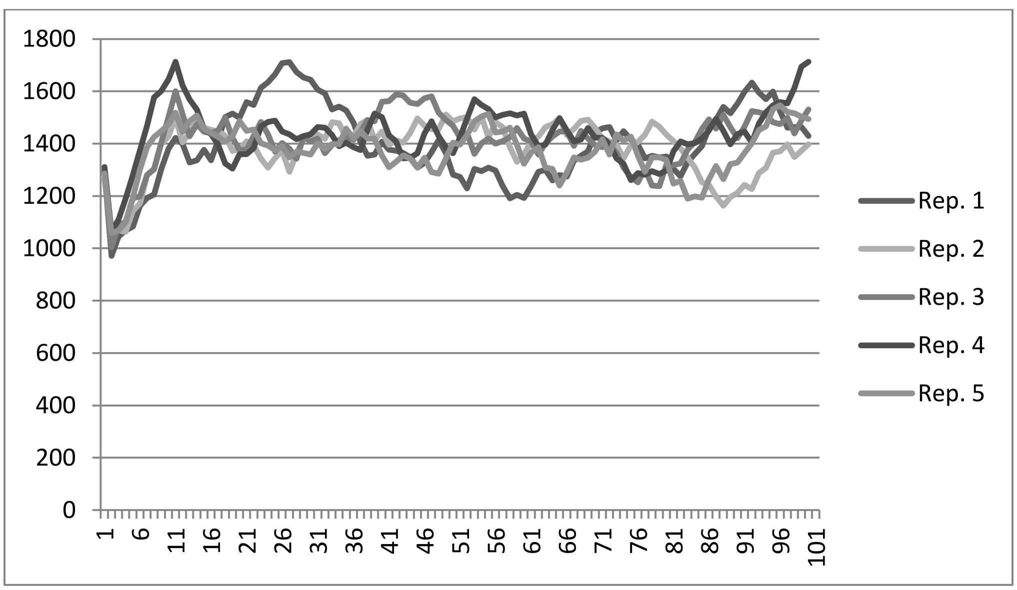

Figure 13: Internal validity. Data trend for population in the network

- ***Sensitivity analysis:*** In order to perform sensitivity analysis for the model, the total number of active links in the network is considered as examination variable. The total number of links in the network is mainly affected by the maximum number of partners a scientist may have in the network. For performing the analysis, three different scenarios are run. In the first scenario, scientists may have only one partner in the network. In the second scenario, the number of possible partners is exactly three partners. And in the third scenario, any number of partners is equally possible. The Figure 14 shows the results for the total number of links during fifty year replications for the three scenarios. All other settings for the three scenarios are the same.

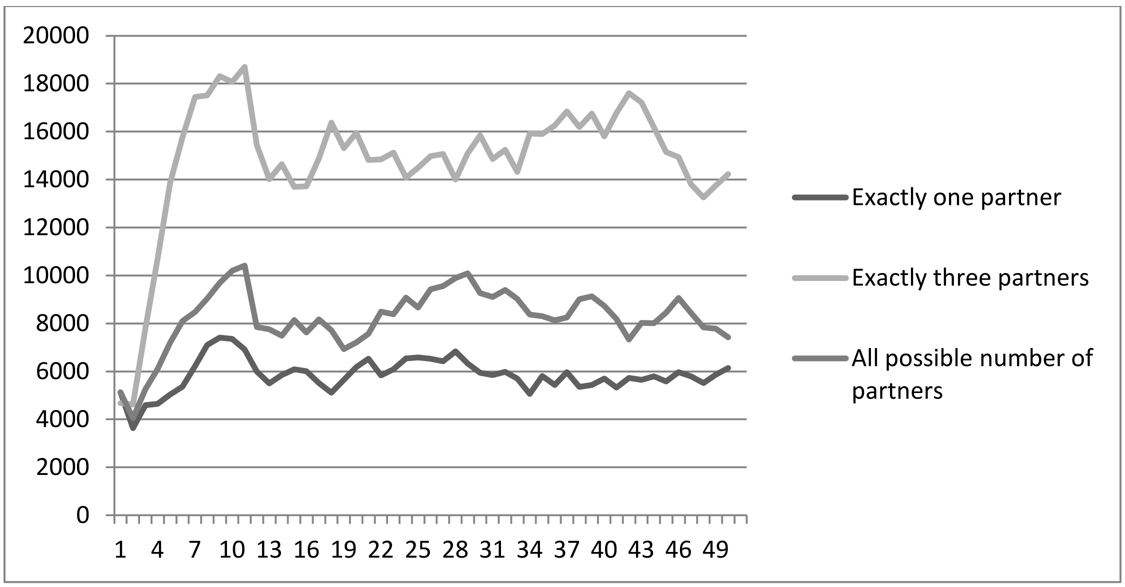

Figure 14: Sensitivity analysis

As the figure shows, the total number of links in the network is sensitive to the possible number of partners. Therefore, the scenario with exactly three partners produces the highest number of links in the network, and the scenario with exactly one partner possibility produces the lowest number of links, during a fifty year replication. These results were expected, since the higher the number of partners, the more links are produced in the network.

- ***Comparison of data trends (Calibration):*** It is necessary to make sure that the simulation model is generating numbers for various variables as intended. For developing the input expressions of the model, ARENA input analyzer have been used as mentioned before. In order to compare the results of the model and that of the database, three variables are chosen and the corresponding developed expression for each is presented in Table 8. The table shows the trend of data from database and ten iterations of the model. The first variable analysed is the number of scientists per cluster, which is actually a direct variable that is fed into model based on the trends in the database. As soon as a new scientist is added to the model, a cluster ID is assigned to it based on the Exponential expression that is fed to the model according to the trend in the database. Since the number generator of Java works randomly, the results for the runs are slightly different. However, the consistency of the results is clear as the expressions of the model are almost the same.

In order to show the fact that the performance of the model represents well the database, the indirect variables of the model need to be compared to the database. The second and third variables shown in Table 8 are indirect variables. The number of the patents per cluster obeys an exponential distribution with mean of 117 in the database. In the model, the number of patents depends on the collaborations among scientists and rate of co-inventorship.

Table 8: Data Trends

Variable	Expression from Database	Expressions from Model				
		Iteration 1	*Iteration 2*	*Iteration 3*	*Iteration 4*	*Iteration 5*
# of Scientists per Cluster	0.5 + EXPO(8.26)	0.5 + EXPO(9.04)	0.5 + EXPO(8.56)	0.5 + EXPO(7.61)	0.5 + EXPO(8.58)	0.5 + EXPO(7.54)
# of Patents per Cluster	2 + EXPO(117)	2 + EXPO(122)	2 + EXPO(117)	2 + EXPO(114)	2 + EXPO(111)	2 + EXPO(115)
# of Collaborations per cluster	10 + EXPO(330)	10 + EXPO(305)	9 + EXPO(331)	10 + EXPO(351)	11 + EXPO(329)	10 + EXPO(345)

Variable	Expression from Database	Expressions from Model *(continue)*				
		Iteration 6	*Iteration 7*	*Iteration 8*	*Iteration 9*	*Iteration 10*
# of Scientists per Cluster	0.5 + EXPO(8.26)	0.5 + EXPO(8.06)	0.5 + EXPO(8.52)	0.5 + EXPO(8.63)	0.5 + EXPO(8.97)	0.5 + EXPO(8.06)
# of Patents per Cluster	2 + EXPO(117)	2 + EXPO(118)	2 + EXPO(115)	2 + EXPO(121)	2 + EXPO(117)	2 + EXPO(117)
# of Collaborations per cluster	10 + EXPO(330)	10 + EXPO(332)	10 + EXPO(344)	9 + EXPO(320)	9 + EXPO(341)	10 + EXPO(335)

As it is illustrated in the table, the number of produced patents in the model is in consistency with the database, since it also obeys exponential trend with slightly different means to those of the database. Finally, the total number of collaborations inside each cluster is analysed. Since the number of collaborations depends on several direct variables in the model (the chooser variables), less consistency is expected in the results. As the table shows, the expression of the trend is still exponential. It is predictable since the number of scientists in the clusters is distributed exponentially. However, the expressions vary from run to run as the decision variables (*chooser variables*) change.

Finally, no model is ever a hundred-percent perfect representation of a system, since all of the complications of the system could seldom be detected. However, while developing a simulation model, the trade off between the accuracy of the model and the cost of increased validation efforts should be considered. As soon as the proper results are attainable from the model and the face validity and assumptions in the model are in an acceptable level, the model is considered to be complete. After the verification and validation of the model is done, it is ready to run various scenarios and analyse the outcomes.

5. Analysis of the model and the results

This part of the thesis deals with the research questions of the thesis and tries to apply various scenarios to the model in order to find answers to these questions. To compare the various scenarios, some basic characteristics of the network must be studied and analysed in common for all the scenarios. As it was also declared in the literature review of the present thesis, the structure of the innovation networks plays the key role in the diffusion of knowledge and production of innovation. The flow of knowledge in the networks depends on some characteristics of the networks, such as degree centralization, betweenness centralization, density, etc.

All the network characteristics are calculated by PAJEK social network analysis software, after it is given the data about each scenario of the model. In order to establish the network by the software, first a vertex is assigned to each distinct identification number of each scientist. Then, the cluster, research field, and organization type of each scientist are given as partitioning conditions in the network. Then for each scientist, its current number of articles plus patent quality and the star status are given as values for each node, indicating the importance of each scientist in the network. Finally, the graph of all the links between pairs of scientists is fed into software in order to analyse the network properties. The age of the links, i.e. the duration of time that the link has been connected between two scientists, indicates the length of acquaintance for scientists and is given to the software as the weight of each link.

5.1 The effects of loyalty on the innovation networks

The cooperative relationships between humans are affected by various criterions. The duration as well as the depth of a relationship is dependent on its purpose and benefits for engaged partners. In a productive academic environment the collaboration ties are created in order to produce new knowledge, so the similarity of goals, required skills of the partner, trust, and records of previous successful collaborations are all very important when scientists choose their partners (Che Mat, Cheung, and Scheepers 2009). Everyday life experience however has shown that there are other personal factors that play a role in the decision procedure of choosing partners, such as culture, family, social life, etc. Therefore, it is expected to see some people changing their partners frequently, while others prefer loyalty to their current partners (Buchan, Croson, and Dawes 2002; Kollock 1994).

Since the process of searching for new partners is time-consuming, people are most of the times willing to remain loyal to their previous partners, even when better choices are available. The loyalty to previous partners causes the structure of collaborative networks to become embedded (Che Mat, Cheung, and Scheepers 2009). It also increases the existence of heterogeneously closed populations (Amaral et al. 2000).

The important factor of an innovation network that is also the main focus of the present thesis is the flow of knowledge between its components (individuals, clusters). The faster and more freely knowledge flows in the network, the faster the innovation grows. The flow of knowledge in the network is closely related to the structure of the network, as in the networks with dynamic links, knowledge has higher chance of transmission (Zander and Kogut 1995). However, in the innovation networks, loyalty to previous partners results in a static network, in which some of the links may never be connected and some routes for flow of the knowledge never shape. Although the survival advantage of loyalty in the network encourages scientists to stick to their previous partners for their new collaborations, the reduced number of potential paths as a result of highly clustered and embedded network structures acts as a disadvantage for the innovation production (Segbroeck et al. 2009).

In a network with strong social ties, there also exists a challenge for new-comers to find collaborative partners, since many of the present scientists have a tendency to work with their current partners, and refuse to start collaborations with new ones. As Kemelgor and Etzkowitz (2001) suggest, becoming a scientist requires knowing other people and be involved in a highly socialized context. They also confirm that having access to professional interactions is crucial for a scientist to survive and improve its position in the network. Without being involved in scientific collaborations and having social support, most of the new scientists may find it difficult to survive in the network, and may remain disconnected from sources of knowledge and isolated consequently.

It can be concluded that loyalty to a previous partner when building new collaborations has an impact on the network structure. In order to analyse the consequences of loyalty in the Canadian innovation network, the duration of collaborative relationships is determined as the indicator of loyalty between partners. The model is built in a way that scientists take into account their previous partnerships while searching for new collaborations. In this regard, there is a higher probability for a scientist to be chosen if there is a record of mutual collaborations in a specific period in the past, as it is illustrated in the methodology section of the present thesis.

In order to analyse the effects of loyalty on the innovation networks, the average age of links between partners is considered as the independent variable in the model. By changing the average age of a link in the network, the duration of loyalty to previous partners is changed in the network, i.e. higher levels of social ties are represented by higher link ages in the network. The simulation model is then run with various settings for the average age of the links, and the behaviour of scientists and the whole network is analysed in order to understand the impact of loyalty on the networks better.

According to the literature, it is expected to have a more embedded network as the average link age between partners increases in the network (Dyer and Singh 1998). Besides, high loyalty amongst scientists decreases the collaborations between existing individuals and new comers, and creation of new collaborative links amongst already existing scientists. This may prevent the connection of separate components in the network, and decrease the growth rate of the largest component proportion in the network. Consequently, there will be more isolated cliques, in which there might be different flows of knowledge. This could increase the probability that some information does not transfer heterogeneously in the network (Tallman et al. 2004).

For analysing the impact of loyalty, different link ages are set as average age of the links between partners from one year of loyalty to thirty years. In the two extreme scenarios, the situations with no loyalty and life-time loyalty are analysed by setting the average age of the links as zero and thirty years (more than the average age of a scientist in the network), respectively. The simulation model is run independently in order to gain the results. The performance of the network is then analysed through charting different indicators of the network structure to compare the results for various scenarios.

5.1.1 Productivity of the network

The first feature of the model which is analysed in order to compare the network under various loyalty levels is productivity. As the Figure 15 shows, the total number of contributions (articles plus patents) remains almost constant in the network. However, as

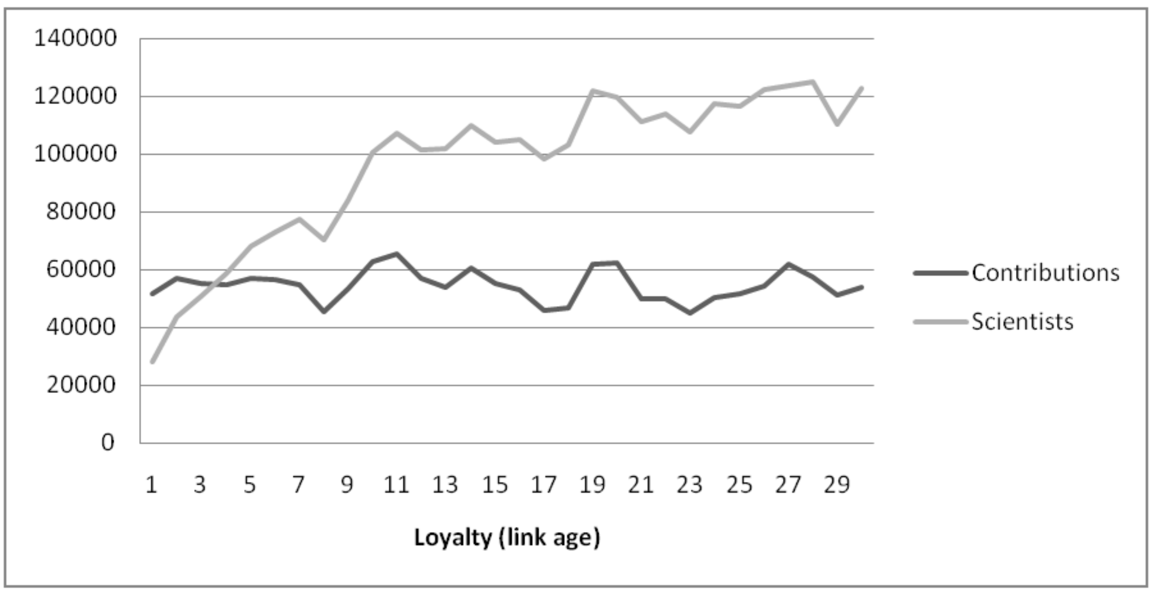

Figure 15: Total Contributions/ Scientists in the Network

the level of loyalty between partners increases, the population of scientists in the network also increases. This implies that the scientists work on average in bigger teams and for the same number of contributions more scientists are required in higher levels of loyalty.

For a better understanding of the network, the first indicator analysed is the innovative productivity of scientists in the network. In order to calculate the productivity of the network, the average total number of publications and patents per scientists is calculated for each scenario. Since the impact of various loyalty levels on the performance of the network needs to be examined, the average productivity of the scientists is a good measure for the overall performance in the network. This indicator is calculated from the following formula for each run of the simulation model separately:

$$Average\ productivity = \frac{\text{Total number of publications} + \text{Total number of patents}}{\text{Population of scientists}}$$

This number is based on the results of fifty-years run of the model, and represents the performance of the network over its total life. The Figure 16 shows the *Average productivity* of the network with thirty different levels of loyalty (i.e. the age of the links between scientists). As the figure illustrates, the overall productivity of the network decreases as the loyalty of scientists to their previous partner increases.

In order to show that the share of the knowledge contribution per scientist decreases as the loyalty level grows, the ratio of star scientists to the whole population of the scientists is presented in the Figure 17. As the figure shows, the number of stars decreases fast as the loyalty increases. This indicates that not only the share of productivity decreases for the network, but also it decreases for highly productive scientists, resulting in less

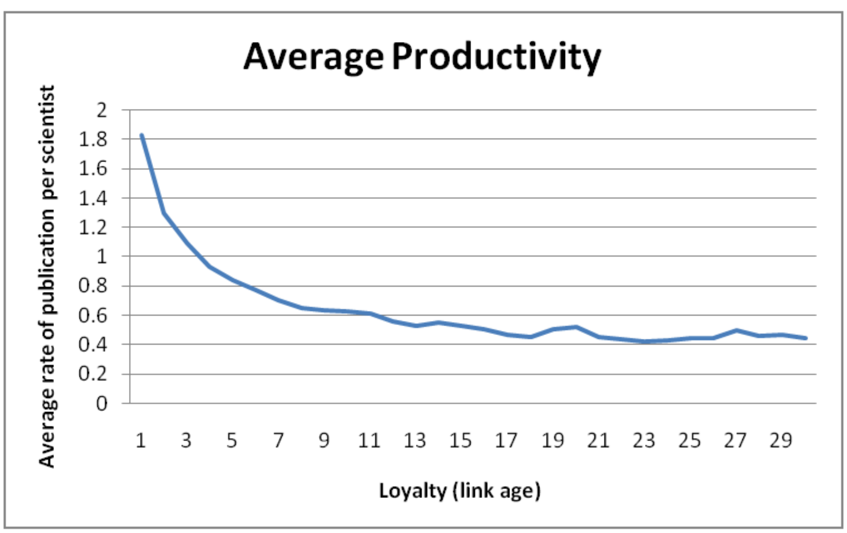

Figure 16: Average Productivity of the Network

number of stars in the network. The figures also suggest that having less loyalty in the network would result in a higher overall productivity.

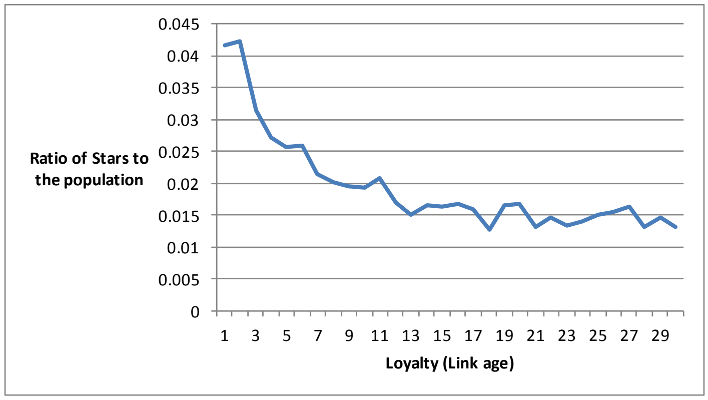

Figure 17: Ratio of Star Scientists to the Whole Population

5.1.2 Network characteristics

In order to analyse the structure of the Canadian biotechnology network for various levels of loyalty levels between scientists, the *degree centralization*[14] and *betweenness centralization*[15] measures of the network nodes are computed. As mentioned before, in the simulation model of the present thesis, the lines connected to each vertex are actually the collaborations between mutual scientists. Since the loyalty in the network is represented by the age of the links, it is expected that the *degree centralization* of the

[14] The degree centrality is a measure of graph theory that indicates the number of lines connected to each vertex. In the simulation model of innovation network, since the vertexes are scientists, this number indicates the number of collaborators for each scientist in the network. Degree centralization is calculated by dividing the variation of nodes' degrees by the highest possible variation in a network of the same size. (De Nooy et al. 2011)

[15] The betweenness centrality indicates the amount of information that passes through each inventor by calculating the proportion of all shortest distances between pairs of other vertices that include this vertex. In other words, the betweenness centrality of a node is defined as the proportion of all shortest paths between pairs of other nodes that contain this node. The variation in the betweenness centrality of nodes in a network is measured by betweenness centralization. (De Nooy et al. 2011)

network increases as the loyalty level increases between scientists. This means that there are more active links connected to a scientist in high loyalty levels.

The corresponding average degree centralization of the network for each loyalty level of the simulation model is extracted from PAJEK, and the results are shown in the Figure 18. As it was expected, the average degree centralization of the network grows, and there is an average of 10 to 11 links per scientist in low levels of loyalty, while in higher levels this average reaches almost 14 links per scientist in the network.

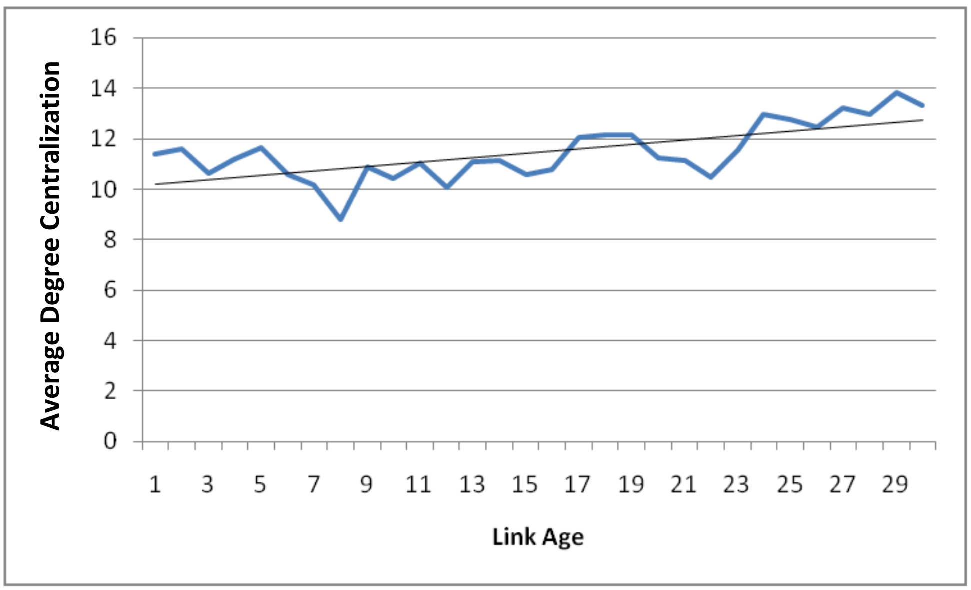

Figure 18: Average Degree Centralization of the Network

Although degree centralization is interpreted in terms of the variance of knowledge flow among the nodes through the network, a higher level of degree centralization does not necessarily result in faster flow of knowledge; i.e. it is possible to have some few nodes with very high degree centralities and all others with low centralities, resulting in a relatively high average.

As another characteristic of the network, changes of betweenness centralization are measured for different levels of loyalty in the network. The results show that the betweenness centralization does not seem to be affected by the loyalty amongst scientists and fluctuates steadily for various scenarios. The Figure 19 shows the betweenness centralization of the simulation experiment for various levels of loyalty.

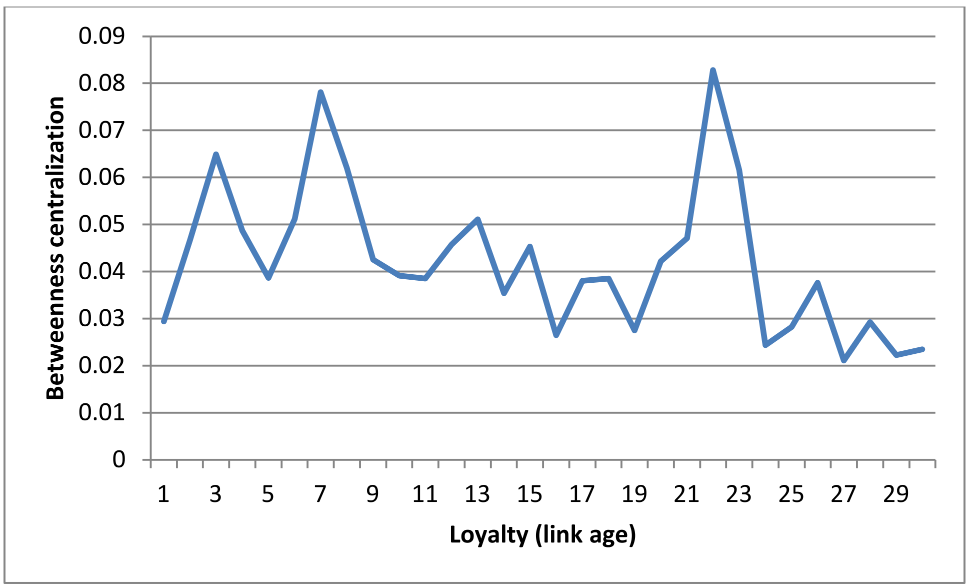

Figure 19: Betweenness centralization of the network

The flow of knowledge in the network is also affected by the density[16] of the network. The higher the density of a network, the faster the knowledge flows between its agents, due to the higher number of connections between them. The density of Canadian

[16] The density of a graph represents the proportion of ties in the graph to the total number of possible ties of a graph with the same number of nodes.

biotechnology network is measured for different loyalty levels, from one year to thirty years, between scientists using PAJEK, and the results are presented in the Figure 20.

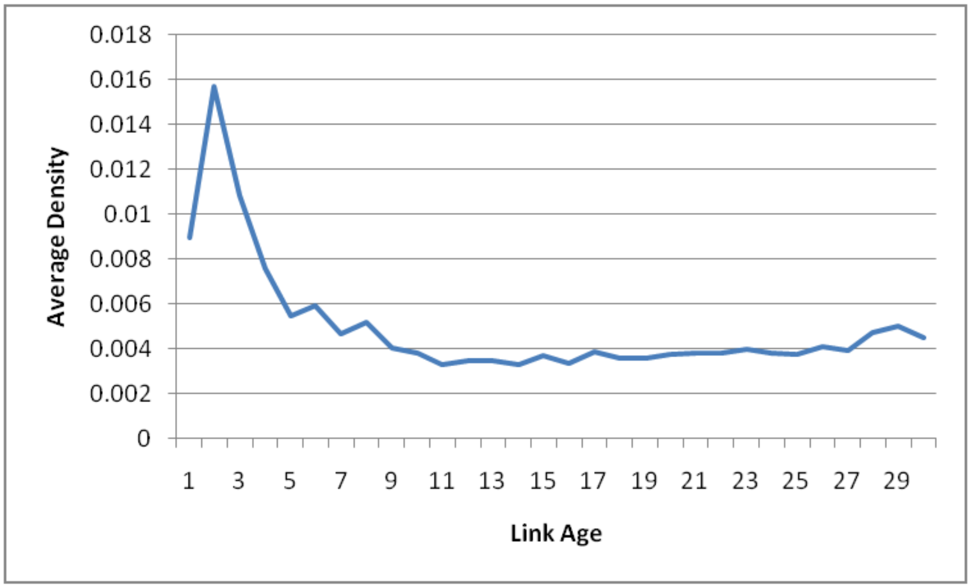

Figure 20: Network Density

As the figure shows, the density of the network decreases fast as the loyalty increases, then the decrease slows down and after reaching the link age of around 8 the density remains relatively stable. This shows that although the growth of loyalty affects the average degree centralization of the network positively, it has a negative impact on the density of the network, resulting in sparser network in higher levels of loyalty.

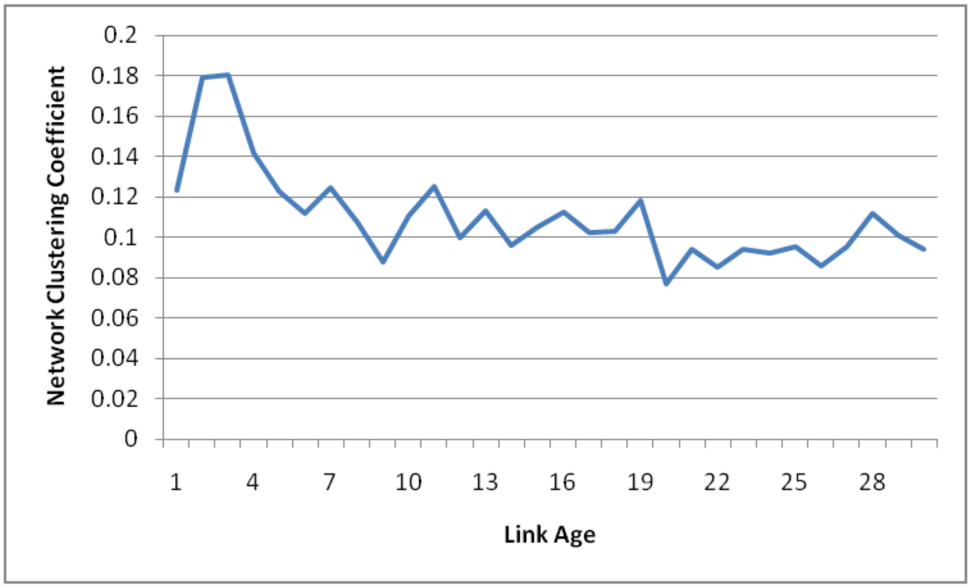

Figure 21: Network Clustering Coefficient

This outcome also confirms that loyalty affects the embeddedness of the network, and reduces the number of potential paths between scientists that could result in more frequent flow of knowledge. Another measure for analysing the embeddedness of the networks is clustering coefficients[17]. Clustering coefficient represents the fraction of collaborators of a node who also collaborate with each other. Networks with higher interconnectivity among their nodes have clustering coefficient closer to one. Therefore, higher clustering coefficient would result in faster transmission of the knowledge amongst the scientists. The clustering coefficients of the network are also measured by PAJEK software for various loyalty levels (link ages). The results are depicted in Figure 21 to show the impact of loyalty on the network average clustering coefficient. As it is seen, the clustering coefficient of the network declines as the collaborative relationships between partners become more long-lasting, i.e. more repetitive. Although the decrease is not very significant (less than 0.05), it still shows the tendency of the network to have a

[17] The clustering coefficient is defined in graph theory as a measure of degree to which nodes in a graph tend to cluster together.

less socialized context. This fact is also undesirable for new scientists, and can affect the performance of new comers in the network negatively, i.e. making it more difficult for them to have access to the main sources of the information in the network.

The last analyzed structural property of the network is the size of the components in the network. As the simulation experiment shows in Figure 22, the proportion of isolated scientists to the total number of components in the network increases as the loyalty level grows among them. This phenomenon most probably happens because in higher levels of loyalty fewer chance of being chosen as a partner for a new scientist is possible, and the chance of having single-scientist innovations increases consequently. This implies that as the loyalty level increases in the network, scientists are more prone to work alone.

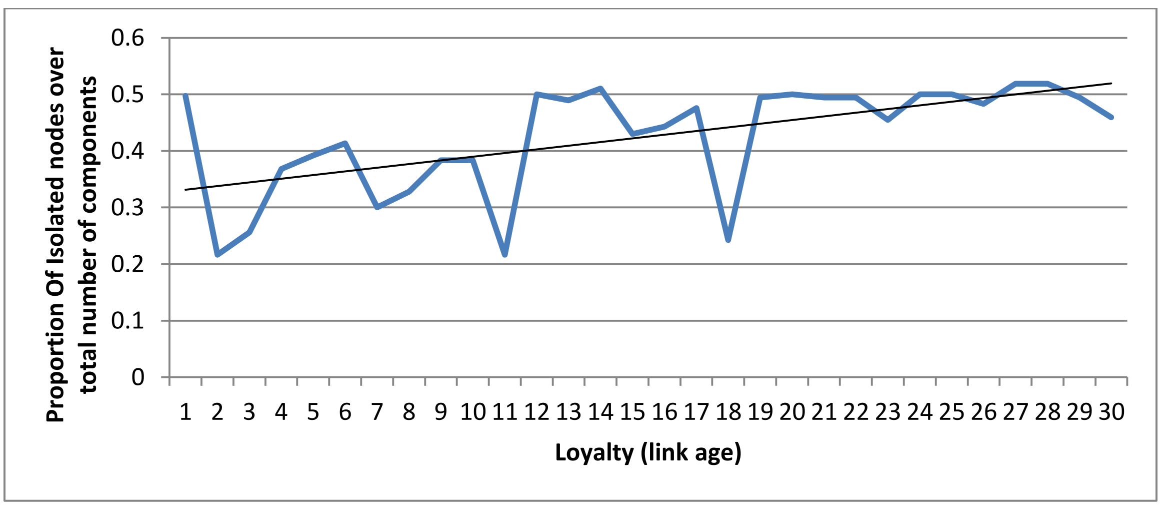

Figure 22: Proportion of isolated scientists to the total number of components

Overall, as the simulation model of this thesis showed, long-lasting collaborative relationships, i.e. repetitive collaborations on the subsequent publications between partners can affect the structure of the innovation networks negatively, resulting in less flow of knowledge among agents of the network. Besides, as the model shows, loyalty results in reduced number of paths among the agents, less productivity of the network and scientists, and affects network characteristics such as centralization, density, and clustering coefficient negatively.

5.2 The role of star scientists in the innovation networks

Since star scientists are the scientists with the extremely high contribution to the scientific output, their role in the innovation networks has been much discussed in the literature (Zucker and Darby 1995; Niosi and Banik 2005; Schiffauerova and Beaudry 2011; Zucker and Darby 1996; Darby and Zucker 2003). The star scientists are the main producers of innovation and knowledge in the network. According to Zucker and Darby (1995) they represent less than one percent of the population of the scientists, but are the authors of more than 15% of the articles and patents.

The literature on biotechnology star scientists reports that, in most cases stars are more likely to repetitively collaborate with the same scientists (Zucker and Darby 1996), which could consequently result in a less socialized network context that reduces the transmission of knowledge to other scientists in the network. Since the stars are the main sources of knowledge in the network, the flow of knowledge might be affected by their presence and absence in the information based innovation networks.

Star scientists usually occupy more central positions in the network (Schiffauerova and Beaudry 2011). Their central position is the result of the higher number of connections of stars than what other scientists have. It is hypothesised that many inventors in the network thus would be isolated in the network if their connection to the stars is lost for any reason (Schiffauerova and Beaudry 2011). However, since as a part of innovation networks, star scientists appear and grow in the networks naturally, the hypotheses regarding their absence from networks cannot be justified by real evidence. In order to analyse the behaviour of innovation networks in the absence of star scientists therefore, a substitution for real world would be required, in which the effects of their absence on the characteristics of the networks could be examined.

It is expected that in a network without star scientists the structure of the network would be more homogeneous, and the links between vertices would be more evenly distributed. Besides, it is expected that also the knowledge production will be more evenly distributed among the scientists if the star scientists are not included in the network.

Buchan, Croson, and Dawes (2002) and later Niosi and Banik (2005) consider "trust" as the basis for scientific collaborations in almost all social contexts, and Kollock (1994) claims there is a relation between trust and reputation in the formation of cooperative and exchange structures. The academic reputation of star scientists hence makes them more trustable than other scientists in the network, which may results in the overemphasis on the trust and the neglect of other factors during the selection of potential collaboration partners.

In order to analyse a network without star scientists, the primary settings of the model are modified in a way that no star scientist exists at the beginning of iterations. The definition of star scientists is also deactivated in the model, and no scientist is thus defined as a star further in the iteration. The characteristics of the innovation network are then analysed and compared for both scenarios, with and without star scientists. In order to gain more even results, five iterations of the model are run for each scenario. The results are then compared for these ten runs of the model.

The next section will numerically analyse the structure of innovation networks, and general characteristics of innovation networks in a scenario without preferred individuals, in which all the scientists have the same value in the network in terms of productivity and innovativeness, and compare the gained results with the current situation of having star scientists in the Canadian innovation network.

5.2.1 Productivity of the network

First, the performance of Canadian biotechnology network is analysed in the presence and absence of star scientists. The average productivity of the scientists (i.e. total number of articles and patents divided by the population of scientists) is measured for both scenarios, and the corresponding results are presented in Figure 23. As it is illustrated in the figure, the average productivity of the network reduces to almost two thirds of its amount in the absence of star scientists in the network. Since the total number of collaborations in the network remains almost the same for the both scenarios (around 57529 collaborations for scenario with stars, and 56308 collaborations for scenario without star scientists), it can be concluded that in the case where no scientists are considered as preferable collaborators, the opportunity for all the scientists to be selected is more similar, and therefore the share of productivity is more evenly distributed among the scientists.

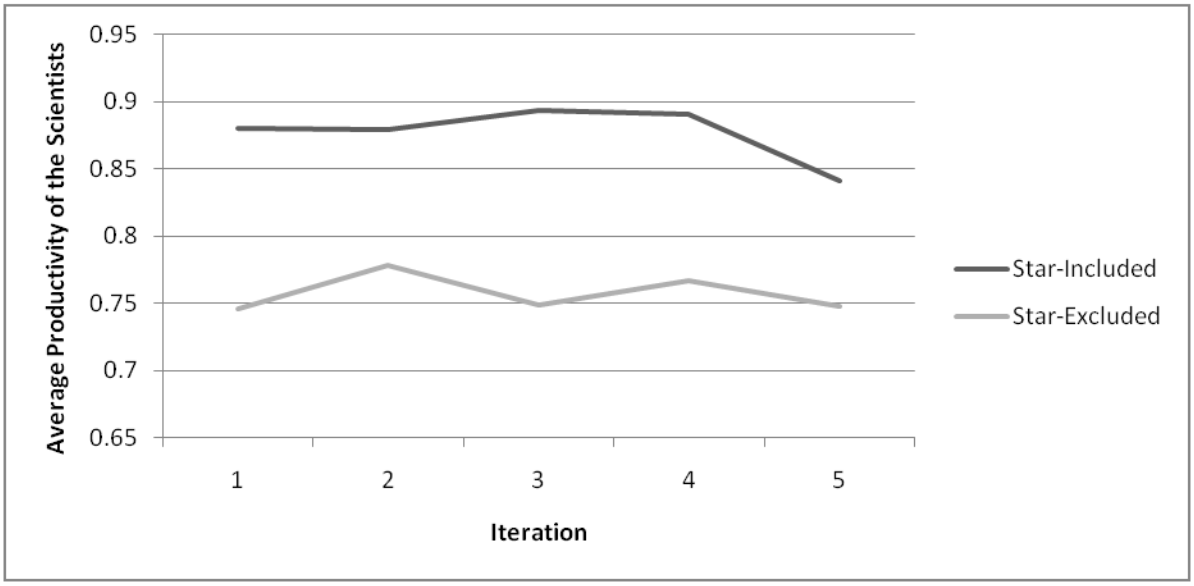

Figure 23: Average Productivity of the Scientists

Figure 24 compares the total number of collaborations in the network. As it is depicted in the figure, the total numbers of links in both scenarios are almost the same. However, the average of total number of links for the five iterations in no-star scenario is slightly smaller than the scenario with stars.

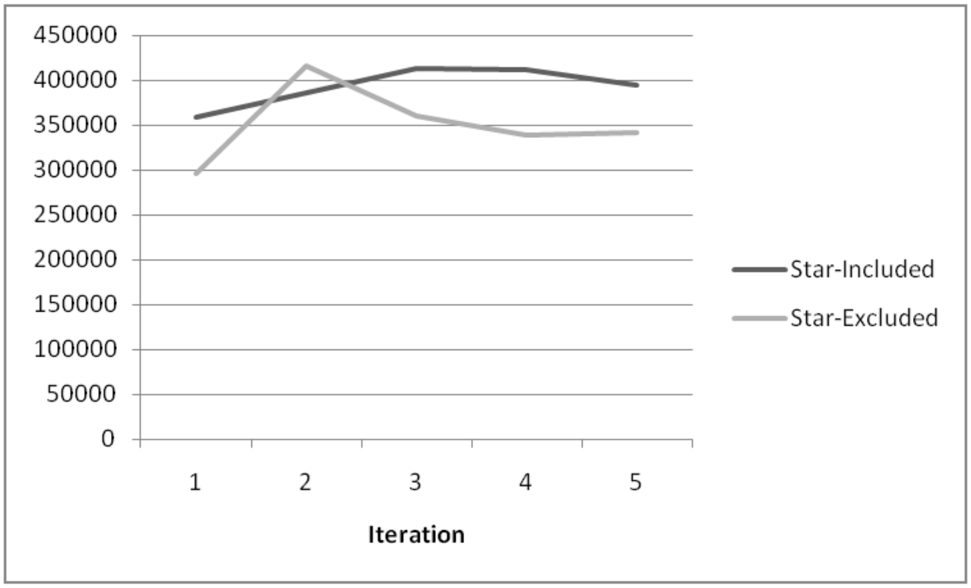

Figure 24: Total Number of links in the Network

In order to analyse the distribution of links in the network in the two scenarios, the average number of links per scientist is compared in Figure 25. As the figure shows, the share of links in the scenario with the star scientists is slightly more than this value in the scenario without the stars. This can imply that in a network without star scientists the links are more evenly distributed, and the possibility of collaboration for scientists is more similar, giving all the scientists the opportunity of being chosen for innovative co-authorships.

The results of the simulation experiment for the two scenarios do not show much variation in the overall number of publications and patents in the network. Generally, as the results of the simulation model show, although the absence of the star scientists does not affect the productivity of the network in terms of total number of patents and publications, it increases the share of productivity per scientist by evenly distributing the collaborations in the network.

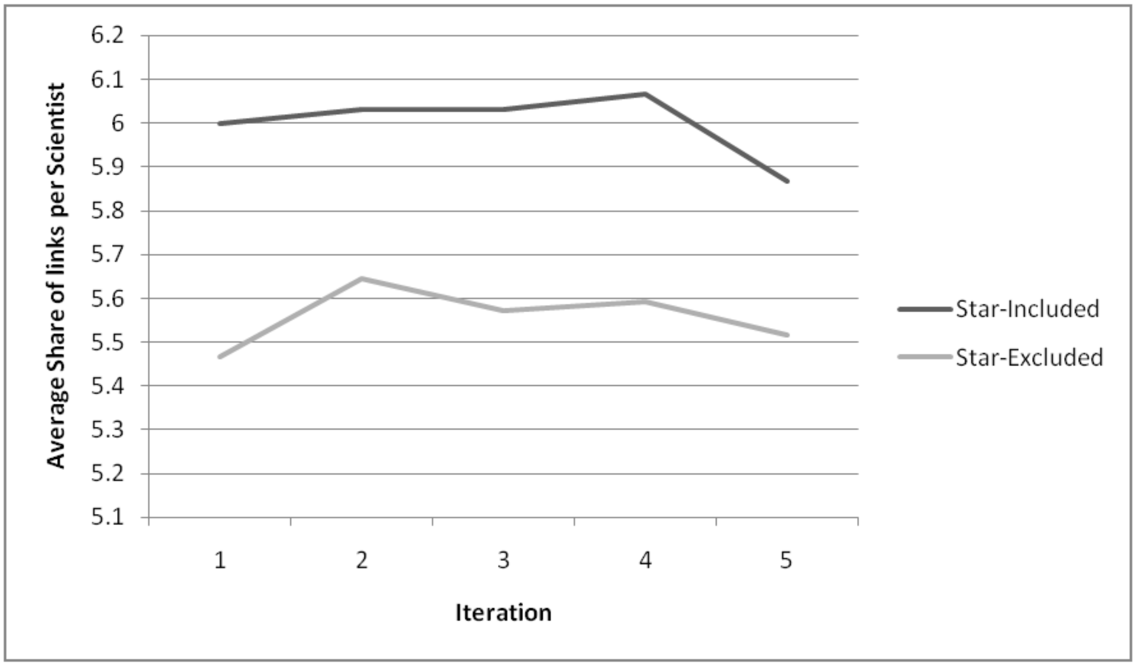

Figure 25: Average Share of links per Scientist

5.2.2 Structural Characteristics of the Network

The structural characteristics of the Canadian biotechnology network are analysed and compared in the presence and absence of star scientists. The results for average betweenness centralization, average degree centralization, average density, average clustering coefficient, and average proportion of isolated nodes to the total number of components are presented in Table 9. The results show no significant change in betweenness centralization and average density of the network for the two scenarios. The betweenness centralization, which refers to the centralization of the nodes in a network and deals with shortest paths in the network, shows no significant change in the presence and absence of stars. However, it should be also taken into consideration that the population of star scientists is very small in comparison with the whole network, and therefore causes no significant structural changes in the two cases.

Table 9: Network Characteristics for the Two Scenarios

	Ave. betweenness centralization	Ave. degree centralization	Ave. density	Ave. clustering coefficient	Ave. (#isolates/#components)
Network with star scientists	0.045	11.362	0.006	0.79	0.849
Network without star scientists	0.045	10.476	0.007	0.85	0.723

Obviously, the total number of links which stars have in the network is significantly higher than in case of other scientists because of the star scientists' more prolific life in the network. However, as the results show, their absence from the network does not cause the other scientists to lose their links, and scientists would still search for new scientists and continue their collaborative actions.

The fact that scientists have a comparable opportunity to find other regular (not necessarily star) scientists and start collaboration for both scenarios is also verified by the unchanged average density in both cases. Since the density represents the proportion of ties in a network relative to the total potential ties, we can conclude that in the network without star scientists, there is still an equal number of links and collaborations in the network, and hence scientists do not have difficulty in finding new partners. Although there is no significant change in the betweenness centralization and density of the network in the absence of star scientists, the Table 9 shows that their absence would result in lower degree centralization, and also higher clustering coefficient in the network.

Average degree centralization, as an indicator of the average number of links per node, has decreased by about 1 link in the network for the second scenario. This was expected since it is assumed that the network in the presence of the stars is more centralized, i.e. more heterogeneous in terms of share of ties between vertices (scientists). The fact that the population of stars is very small (about 5% of the network) makes this relatively significant change even more considerable, implying that the share of ties among other scientists is very low in amount. Obviously, the flow of knowledge in the network is very much affected by the share of ties between the nodes in the graph and mathematically a network with higher average degree centralization has potentially more flow of knowledge among its nodes. The results of the simulation exercise so far suggest that star scientists are not only productive scientists by themselves, but also can improve the knowledge transmission performance of the whole network by making the network more centralized.

The next network characteristic analyzed for the two scenarios is clustering coefficient of the network. The Table 9 shows that the clustering coefficient of the network in the absence of the starts is closer to one, indicating that in a network without stars there is more interconnectivity between the nodes than in a network including them. As it was mentioned before, clustering coefficient measures the proportion of partners of a scientist who also have collaboration with each other. According to the results, there is 85 percent chance for two individuals with a common collaborator to also have partnership together in a star-excluded network, whereas this chance is around 79 percent in a star-included scenario.

This suggests that a network with stars is less interconnected and has more hubs than a network in which all the scientists have the same possibility of access to the knowledge, regardless of their previous performance in the network. This also provides an advantage for new scientists, who might get completely disconnected in the network with low interconnectivity measure.

In order to compare the flow of knowledge in the network, the number of isolated nodes is determined for each scenario, and the proportion of isolates to the total number of components in the network are presented in the Table 9. According to the table, there is a higher chance of having isolates in a star-included network. Eventually, we can say that new comers in a network without star scientists have higher possibility of survival, since there is a higher chance of having access to new knowledge and getting involved in innovative collaborations for them.

Overall, as it can be concluded from simulation experiment results, the presence of stars has positive impact on the average productivity per scientist in the network. The overall network characteristics are comparable for the two scenarios. However, in a network without stars, there exists more interconnectivity and consequently better flow of knowledge amongst scientists, which would result in faster growth of innovation, and also there is higher chance for new scientists to be selected as a partner, resulting in less isolated components in the network.

5.3 The role of Gatekeepers in the innovation networks

As an interpersonal and information centered process, the creation of innovation is the key to the emergence of communication networks and various roles for scientists acting as the networks' agents (Keller 1991). One of the critical roles in the network, the role of Gatekeepers has been introduced in previous parts of the thesis.

Sosa and Gero (2005) define Gatekeepers as individuals with access to outside sources of knowledge, who maintain the role of intermediaries in the dissemination of knowledge by having personal communications inside and outside their clusters. As it is also mentioned in the literature review section, no other research study has examined and compared the network dynamics in case of the presence and the absence of Gatekeepers.

Since Gatekeepers are defined in this thesis as individual inventors in the system who have several communicational links inside and outside their clusters, it is obvious that if the Gatekeepers are completely lost, i.e. most of the outside cluster links disappear, the whole network would be split into a number of local clusters working independently, and the components would be more isolated. Heikkinen et al. (2007) state that Gatekeepers not only have a significant impact on the success of the networks, but also improve the performance of individuals connected to them in the network. They argue that the presence of Gatekeepers in the networks is inevitable, since they have emerged naturally and remain in the network by natural requirements and connections that take place in the network. The critical role of Gatekeepers as the possessors of the resources and controllers of the connections can even affect the direction of the research in the network.

As it was discussed before, establishing any new collaborative link requires time, trust and sometimes money. Therefore, a Gatekeeper becomes more and more valuable in the network as the number of his/her connections grows, both inside and outside the cluster. Although the presence of Gatekeepers is vital for the networks as the directors of information and opportunities, the scientists with no direct connection to Gatekeepers in the network are in danger of being isolated in the network because of the low information flows connected to them. This section will examine the structure of the innovation networks and the flow of information in case that all the scientists in the network have the same opportunity for making external connections outside their clusters.

In a static network with invariable Gatekeepers, the social ties are tighter, and there would be a smaller possibility for new scientists of making connections in the network. In the present part of the thesis two scenarios will be compared in order to analyse the structure of the networks with and without constant Gatekeepers. In the first scenario, the Gatekeeper status of a scientist is set as valuable in the choosing procedure of partners, which results in the survival of the Gatekeepers, while in the second scenario the Gatekeepers have the same probability of being chosen as other scientists, which makes the same opportunity for all the scientists in the network to become Gatekeepers.

5.3.1 Productivity of the network

In the network where only some distinctive scientists are responsible for the inflow of knowledge from outside geographical clusters, fewer scientists may try to construct their own links outside their clusters and merely depend on Gatekeepers in this regard.

Therefore, in case of having no distinct Gatekeeper, it is expected to have a greater number of links per scientist in the network, providing individual inflows of knowledge to the cluster.

The results of five separate iterations of the simulation model for Gatekeeper-included and Gatekeeper-excluded scenarios show that there is a higher total share of external and internal links per scientists for the Gatekeeper-excluded scenario (Figure 26: Average Share of Links per Scientist). The results illustrate that the average of total number of links per scientist is 5.9 in a Gatekeeper-included network, while this number increases to 6.4 in a Gatekeeper-excluded network. Therefore, the results of the simulation experiment have confirmed our hypothesis of having individual scientists building up their own links in the network when no one is responsible for the inflow of the knowledge to the cluster.

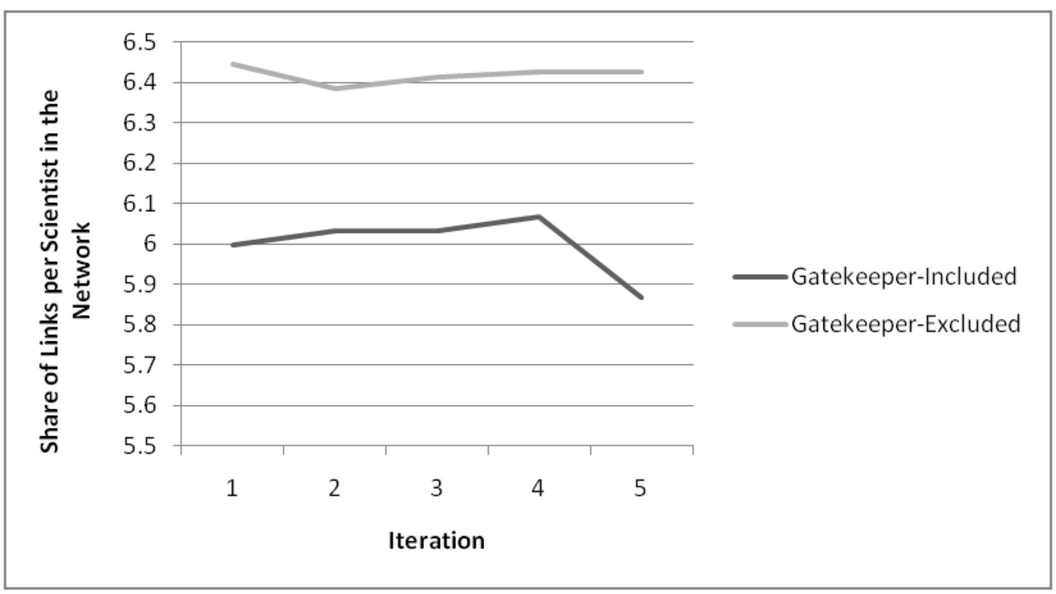

Figure 26: Average Share of Links per Scientist

The average productivity per scientist is depicted in the Figure 27, showing that the average individual productivity in the network in the presence of Gatekeepers is almost doubled. At the first glance, it may seem a pure advantage for the network to have static Gatekeepers (Gatekeepers who are recognized and do not change over time), since it results in higher average productivity of the scientists in terms of their patenting and publishing activities. However, taking a closer look at the total number of contributions (patents plus articles) and total population of scientists, as the Figure 28 A and B, reveals that in the both scenarios the total number of contributions are almost the same, while the Gatekeeper-excluded scenario contains twice bigger population of scientists, resulting in lower share of the knowledge production per scientist in the network.

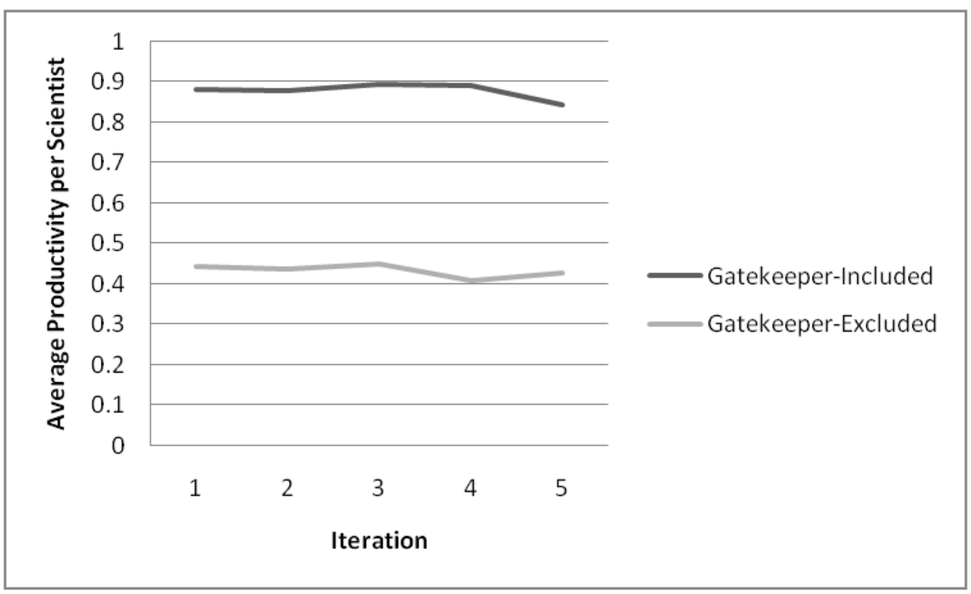

Figure 27: Average Productivity per Scientist in the Network

This comparison implies that, the outcome of a network in terms of total innovative contributions remains almost unchanged between the two scenarios, but the chance of survival for a scientist is much higher in a Gatekeeper-excluded network. The reason is that in a network where all the scientists have the same opportunity of making connections outside their clusters, the chance of survival for a scientist is higher (a scientist who cannot be productive for more than one year ceases to exist in the model).

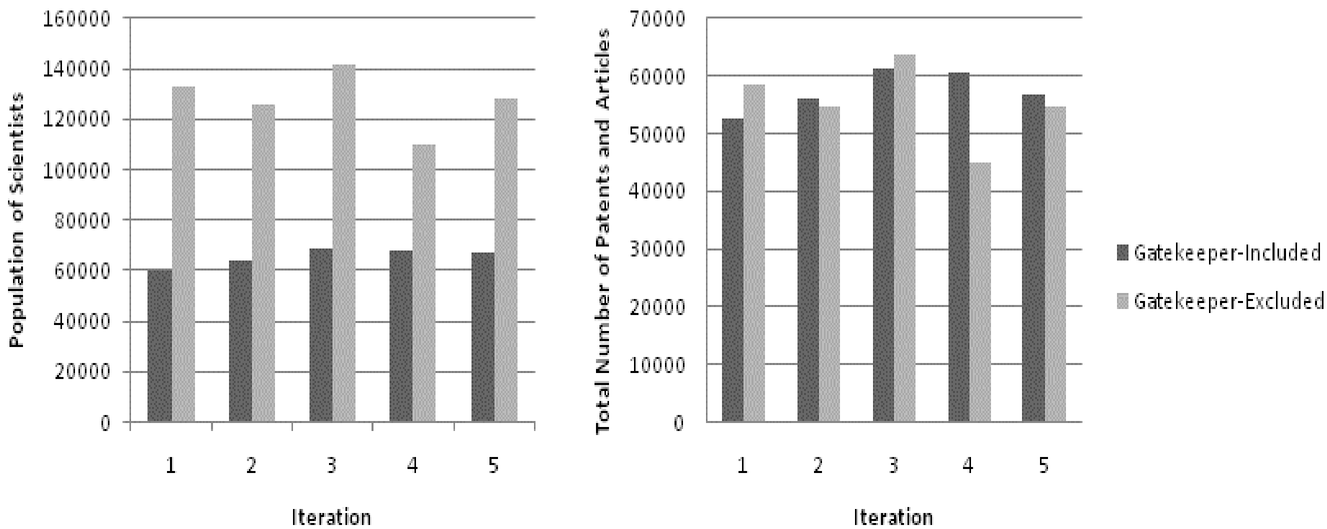

Figure 28: (A-Left) Total Number of Scientists (B-Right) Total Number of Publications and Patents

The comparison of productivity for the two scenarios proves that, although more scientists find the opportunity of getting involved in innovative collaborations when no Gatekeepers are recognized, the total quantity of the patents and articles in the network does not increase. Besides, since the previous experience of a scientist is hypothesised to have positive impact on his/her chance to be included in further collaborations, a network with a lower average productivity per scientist is also expected to have less patents and publications. However, this cannot be confirmed with the present results and some further studies are required in this regard.

5.3.2 Structural Characteristics of the Network

The results of the simulation experiment of the thesis have been fed to the PAJEK software in order to calculate the network characteristics of the two scenarios including and excluding Gatekeepers. These are presented in the Table 10. The Table 10 shows the average values for four structural characteristics (betweenness centralization, degree centralization, density, clustering coefficient, and proportion of isolates to the total number of components) that are computed for five iterations of the model for the two scenarios separately.

Table 10: Network Characteristics

	Ave. betweenness centralization	Ave. degree centralization	Ave. density	Ave. clustering coefficient	Ave. (#isolates/#components)
Gatekeeper-included	0.045	11.362	0.006	0.79	0.849
Gatekeeper-Excluded	0.017	13.521	0.004	0.77	0.941

The average betweenness centralization in the network reduces greatly (from 0.045 to 0.017) in the absence of the Gatekeepers in the network, as it is shown in the table. Since the betweenness centralization represents variation among betweenness centralities of the scientists, this reduction is predictable. The presence of the Gatekeepers as the intermediaries of the flow of knowledge in the network results in higher betweenness centralities for them in comparison with other scientists. This happens because the

Gatekeepers shape hub-shaped connections in the network and play as the center of the hubs, having shortest paths running through them than other individual scientists.

The innovation network benefits from the presence of the Gatekeepers in a way that with less number of ties between scientists, more transmission of knowledge is generated. This also happens because the presence of the Gatekeepers provides the network with more static links overtime. The Gatekeepers in the network act as static bridges over long distances between clusters. Therefore, a Gatekeeper-excluded network, where each scientist tries to build up his/her own ties to gain access to the external knowledge, inevitably results in more redundant and dynamic ties in the network.

The higher average degree centralization of the network nodes also confirms that the average number of links per scientist increases in the Gatekeeper-excluded network. The fact that a Gatekeeper-included network has higher average betweenness centralization and lower average degree centralization could be interpreted by considering the role of the Gatekeepers in the network. The Gatekeepers, as the intermediaries of the knowledge exchange between clusters act as a hub and reduce the total number of between-cluster links in a network, resulting in a more static network. This also clarifies why a Gatekeeper-excluded network exhibits higher average degree centralization, which is an indicator of the total number of ties that are incident upon a node in a graph.

The results also show that the average density of the network is higher for the Gatekeeper-included network, which could result in faster flow of knowledge between nodes of the network. Besides, the clustering coefficient of the network is also higher when Gatekeepers are recognized and static in the network. However, this difference is not as significant as the difference in average betweenness centralization and average degree centralization, the two measures that deal with the average centralization of vertices in the network.

As the last indicator, the proportion of the number of isolated components to the total number of components is calculated for the both scenarios over five runs of the model. As Table 10 shows, there is a lower number of isolate scientists in the Gatekeeper included network. This also supports the idea that the presence of the Gatekeepers improves the connectivity of the clusters in the network and prevents individuals, especially new comers, from becoming detached from the flow of the knowledge in the network.

Overall, the results of the simulation study suggest that the presence of the Gatekeepers in the innovation networks benefits the network significantly, both in terms of productivity and network characteristics that lead to more flow of knowledge through the network. It can be also concluded that most of the scientists from other geographical clusters prefer to bridge through intermediaries rather than directly connecting to scientists from other clusters.

6. Conclusions; Limitations; and Future Study

6.1 Conclusions

This project has been the first attempt to address the role of individual scientists in the Canadian biotechnology innovation network using simulation modeling. The characteristics of the innovation networks have been of interest to the researchers for decades, and there is a vast literature on the organizational-level of study in this regard. The goal of the present thesis has been to shed some light on the individual-level of study of the innovation networks, and propose the simulation modeling as a suitable methodology for this type of analysis.

In doing so, this project first investigates the possibility of the use of two existing available databases of articles and patents in the Canadian biotechnology network. Then an agent-based simulation model of the scientists has been developed for the experimentation with the variables and structural properties of the network. The simulation model has been developed using JAVA programming language, and has been made as flexible as possible for future analyses and studies. According to the databases, some factors have been detected as decisive factors for the scientists in the procedure of establishing new collaborations, such as geographical proximity, similarity of fields of study, organization types, previous background of collaborations, and the role of scientists as Gatekeepers or stars.

Apart from the development of the simulation model of the individual-level of innovation networks, there are three research questions which have been raised, and the simulation experiment has been employed to find answers to these questions by proposing various scenarios of the behaviour of the model.

The first series of scenarios examines the effect of the repetitiveness of the collaborative relationships among scientists on the overall efficiency of the innovation network in terms of its innovative productivity and its knowledge transmission capability. In order to investigate this, the simulation experiment has been run for thirty different durations of loyalty to previous partners. The corresponding results showed that the average share of productivity per scientist declines as the social ties among loyal partners become stronger in the network. Besides, the structural properties of the innovation network that account for knowledge transmission capabilities, such as betweenness centralization, degree centralization, and density, showed negative results of long-lasting levels of collaborations.

The second research question of the thesis deals with the role of highly prolific scientists, known as stars, on the productivity, and structural characteristics of the network. In order to find answer to the question, two scenarios which consider either the presence or the absence of the stars in the network have been experimented by the simulation model. The results showed that the rate of innovative productivity per scientist is lower in a star-excluded scenario. However, the total number of collaborative activities is higher in a star-included scenario. The simulation experiment displayed positive overall influence of stars on the innovative activities of the network, but on the other hand, it showed reduction in the interconnectivity of the scientists in the presence of stars, that could have negative consequences on the flow of knowledge in the network.

The last research question of the thesis examines the role of Gatekeepers as the intermediaries of the knowledge exchange in the networks. Two scenarios, i.e. the network in the presence and in the absence of the Gatekeepers, are experimented by the simulation model. Based on the results, it can be concluded t that the presence of the Gatekeepers in the network enhances both the productivity performance, and the structural properties of the network.

6.2 Limitations

Several limitations were encountered during the analysis of the present thesis.

First, the structure of data in the two databases was not completely compatible. This resulted in slight mismatches of the records, and might have resulted in some bias in the results.

Besides, not all the relationships among scientists could have been captured from the databases, since some of the records had incomplete data. However, the total number of incomplete records was negligible. Moreover, there was a lack of information on the collaboration durations in the databases, which was solved by making several assumptions in the model.

Next, since the two databases only reflected the successful scientific collaborations, i.e. only the articles and patents that have been published or registered, there must have been a certain amount of collaborative relationships which has been neglected because they never resulted in any final published outcome. On the same note, we are aware that most of the formal and informal relationships are in fact never recorded.

Moreover, since the two databases included only the recent collaborations in Canadian biotechnology network, i.e. from 1952 to 2006, many other possible collaborative relations either before this period, in other sectors rather than biotechnology, or in other regions rather that Canada might have been neglected.

The fact that there was only one series of data available also confined the verification and validation steps of the model, i.e. the model was validated using the same data that was used to generate it. However, it was also inevitable since no other source of data with the same properties is available. In addition, the current databases could not have been divided into two sections, since all the correlations were interconnected and dependant, both geographically and chronologically.

Finally, the programming software used for the present project is not the best of its kind, and this resulted in slow runs of the model for each scenario. However, the same methodology could be used in the more appropriate software for the future similar analyses.

6.3 Recommendations

Finally, this thesis proposes some recommendations that could be taken into consideration for the future studies.

First, it is recommended to include other critical factors such as funding, collaboration costs, and collaboration durations in the modeling for the future analyses in order to have a more realistic model. The drawback effect of cost of establishing new connections can be used to perform optimization research and propose applicable regulations and policies for the scientists, as well as organizations and government in order to facilitate the circulation of knowledge and improve the innovation production speed in the networks.

Next, it is recommended to use different databases of various regions to create the networks in order to be able to validate the model more effectively, and make the simulation more applicable to other innovation networks of different fields of science. Moreover, the information on the citations of the articles can be extracted and used as an evaluator of the quality of the articles, in the future research.

The geographical regions, organizational types, and fields of study could also be analysed in more detail with the appropriate data, if available, in order to obtain more specific results. Finally, in order to improve the choosing procedure of the collaborative partners a better understanding of the factors based on which the scientists make their selection decisions is needed. This could be achieved by developing a questionnaire and conducting a survey, which can shed some light on the intentions of the scientists.

References

Acs, Z. J & Audretsch, D. B, 1988. Innovation in large and small firms: an empirical analysis. The American Economic Review, 78(4).

Acs, Zoltan J. & Audretsch, David B., 1987. Innovation in large and small firms. Economics Letters, 23(1).

Aharonson, B.S., Baum, J.A.. & Feldman, M.P., 2004. Industrial clustering and the returns to inventive activity: Canadian biotechnology firms, 1991-2000. Danish Research Unit for Industrial Dynamics Working Paper, pp.04–03.

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

Amaral, L.A.N. et al., 2000. Classes of small-world networks. Proceedings of the National Academy of Sciences, 97(21), pp.11149 -11152.

Banks, J. & Carson, J.S., 1984. Discrete-event system simulation, Prentice-Hall.

Baptista, R. & Swann, G.M.P., 1999. A comparison of clustering dynamics in the US and UK computer industries. Journal of Evolutionary Economics, 9(3), pp.373-399.

Baptista, R. & Swann, P., 1998. Do firms in clusters innovate more? Research Policy, 27(5), pp.525-540.

Beaucage, J.S. & Beaudry, C., 2006. The Importance of Knowledge Networks within Canadian Biotechnology Clusters. In International Schumpeter Conference, Sophia-Antipolis.

Becheikh, N. & Su, Z., 2005. L’organisation virtuelle: un avenir qui se dessine. Revue française de gestion, 31(154), pp.93-110.

Becheikh, N., Landry, R. & Amara, N., 2006. Lessons from innovation empirical studies in the manufacturing sector: A systematic review of the literature from 1993-2003. Technovation, 26(5-6), pp.644-664.

Buchan, N.R., Croson, R.T.. & Dawes, R.M., 2002. Swift neighbors and persistent strangers: A cross-cultural investigation of trust and reciprocity in social exchange. ajs, 108(1), pp.168–206.

Burt, R., 1992. Structural holes: the social structure of competition (Harvard, MA, Harvard University Press).

Burt, R.S., 1980. Autonomy in a social topology. American Journal of Sociology, pp.892–925.

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

Carneiro, A., 2001. The role of intelligent resources in knowledge management. Journal of Knowledge Management, 5(4), pp.358-367.

Che Mat, N.A., Cheung, Y. & Scheepers, H., 2009. Partner Selection: Criteria for Successful Collaborative Network.

Cohen, W.M. & Levinthal, D.A., 1990. Absorptive capacity: a new perspective on learning and innovation. Administrative science quarterly, 35(1).

Coombs, R., Narandren, P. & Richards, A., 1996. A literature-based innovation output indicator. Research Policy, 25(3), pp.403–413.

Cowan, Robin & Jonard, Nicolas, 2004. Network structure and the diffusion of knowledge. Journal of Economic Dynamics and Control, 28(8), pp.1557-1575.

Cowan, R. & Jonard, N., 2003. The dynamics of collective invention. Journal of Economic Behavior & Organization, 52(4), pp.513-532.

Dahl, M.S. & Pedersen, C.Ø.R., 2004. Knowledge flows through informal contacts in industrial clusters: myth or reality? Research Policy, 33(10), pp.1673-1686.

Damanpour, Fariborz, 1987. The Adoption of Technological, Administrative, and Ancillary Innovations: Impact of Organizational Factors. Journal of Management, 13(4), pp.675 -688.

Damanpour, Fariborz & Evan, W.M., 1984. Organizational Innovation and Performance: The Problem of "Organizational Lag." Administrative Science Quarterly, 29(3), pp.392-409.

Darby, Michael R. & Zucker, Lynne G., 2003. Grilichesian Breakthroughs: Inventions of Methods of Inventing and Firm Entry in Nanotechnology. National Bureau of Economic Research Working Paper Series, No. 9825. Available at: http://www.nber.org/papers/w9825 [Accessed June 23, 2011].

VAN DE VEN, A.H. & Rogers, E.M., 1988. Innovations and Organizations. Communication Research, 15(5), pp.632 -651.

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

Dyer, J.H. & Singh, Harbir, 1998. The Relational View: Cooperative Strategy and Sources of Interorganizational Competitive Advantage. The Academy of Management Review, 23(4), pp.660-679.

Ettlie, J.E., 1980. Adequacy of stage models for decisions on adoption of innovation. Psychological reports, 46(3), pp.991–995.

Fári, M.G., Bud, R. & Kralovánszky, P.U., 2001. The History of the Term Biotechnology: Károly Ereky and His Contribution. In presentation at the Fourth Congress of Redbio–Encuentro Latinoamericano de Biotecnologia Vegetal, Goiânia, Brazil, June. pp. 4–8.

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

Fleming, L., King, C. & Juda, A.I., 2007. Small Worlds and Regional Innovation. ORGANIZATION SCIENCE, 18(6), pp.938-954.

Fortino, G., Garro, A. & Russo, W., 2005. A discrete-event simulation framework for the validation of agent-based and multi-agent systems. In Proceedings of the Workshop on Objects and Agents (WOA), Camerino. pp. 14–16.

Frenken, Koen, 2000. A complexity approach to innovation networks. The case of the aircraft industry (1909-1997). Research Policy, 29(2), pp.257-272.

Frenken, K. et al., 2009. Death of Distance in Science? A Gravity Approach to Research Collaboration. Innovation Networks, pp.43–57.

Galende, J. & de la Fuente, J.M., 2003. Internal factors determining a firm’s innovative behaviour. Research Policy, 32(5), pp.715-736.

Gauvin, S., 1995. Networks of innovators: Evidence from Canadian patents. Group Decision and Negotiation, 4(5), pp.411–428.

Gemünden, H.G., Ritter, T. & Heydebreck, P., 1996. Network configuration and innovation success: An empirical analysis in German high-tech industries. International Journal of Research in Marketing, 13(5), pp.449-462.

Gilbert, N., 2004. Agent-based social simulation: dealing with complexity. The Complex Systems Network of Excellence.

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

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

Gopalakrishnan, S. & Damanpour, F., 1997. A review of innovation research in economics, sociology and technology management. Omega, 25(1), pp.15-28.

Gould, R. & Fernandez, R., 1989. Structures of Mediation: A Formal Approach to Brokerage in Transaction Networks. Sociological Methodology, 19, pp.89-126.

Guardiola, X. et al., 2002. Modeling diffusion of innovations in a social network. Physical Review E, 66(2), p.026121.

Hanneman, R.A. & Riddle, M., 2005. Introduction to social network methods, University of California Riverside.

Hargadon, A. & Sutton, R.I., 1997. Technology Brokering and Innovation in a Product Development Firm. Administrative Science Quarterly, 42(4), pp.716-749.

Harrisson, D. & Laberge, M., 2002. Innovation, Identities and Resistance: The Social Construction of An Innovation Network. Journal of Management Studies, 39(4), pp.497-521.

Heikkinen, M.T. et al., 2007. Roles for managing in mobile service development nets. Industrial Marketing Management, 36(7), pp.909-925.

Henderson, R. & Cockburn, I., 1996. Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery. The RAND Journal of Economics, 27(1), pp.32-59.

Howells, J., 2006. Intermediation and the role of intermediaries in innovation. Research Policy, 35(5), pp.715-728.

Kale, P., Singh, H. & Perlmutter, H., 2000. Learning and protection of proprietary assets in strategic alliances: Building relational capital.

Keller, R.T., 1991. Gatekeeper communication networks and technological innovation: A study of U.S. and Mexican R&D organizations. The Journal of High Technology Management Research, 2(1), pp.1-13.

Kemelgor, C. & Etzkowitz, H., 2001. Overcoming isolation: Women's dilemmas in American academic science. Minerva, 39(2), pp.153–174.

Kim, Y. & Vonortas, N.S., 2005. Technology licensing partners. Journal of Economics and Business, 58(4), pp.273-289.

Kimberly, J.R. & Evanisko, M.J., 1981. Organizational Innovation: The Influence of Individual, Organizational, and Contextual Factors on Hospital Adoption of Technological and Administrative Innovations. The Academy of Management Journal, 24(4), pp.689-713.

Kl\ügl, F., 2008. A validation methodology for agent-based simulations. In Proceedings of the 2008 ACM symposium on Applied computing. pp. 39–43.

Kollock, P., 1994. The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment, and Trust. American Journal of Sociology, 100(2), pp.313-345.

Krugman, P.R., 1991. Geography and trade, MIT Press.

Laursen, K. & Foss, N.J., 2003. New human resource management practices, complementarities and the impact on innovation performance. Cambridge Journal of Economics, 27(2), pp.243 -263.

Lenz-Cesar, F. & Heshmati, A., 2010. Agent-based Simulation of Cooperative Innovation. Discussion Papers.

Ma, T. & Nakamori, Y., 2005. Agent-based modeling on technological innovation as an evolutionary process. European Journal of Operational Research, 166(3), pp.741-755.

Macdonald, S. & Williams, C., 1993. Beyond the boundary: An information perspective on the role of the Gatekeeper in the organization. Journal of Product Innovation Management, 10(5), pp.417-427.

Macdonald, S. & Williams, C., 1994. The survival of the Gatekeeper. Research Policy, 23(2), pp.123-132.

Marsden, P.V., 1982. Brokerage behavior in restricted exchange networks. Social structure and network analysis, 7(4), pp.341–410.

Marshall, A., 1961. Principles of economics, 1890. London: McMillan.

Martin, R. & Sunley, P., 2003. Deconstructing clusters: chaotic concept or policy panacea? Journal of Economic Geography, 3(1), pp.5 -35.

Mayer, K.J. & Argyres, N.S., 2004. Learning to contract: Evidence from the personal computer industry. organization Science, pp.394–410.

McMillan, G.S., Narin, F. & Deeds, D.L., 2000. An analysis of the critical role of public science in innovation: the case of biotechnology. Research Policy, 29(1), pp.1–8.

Meyer-Stamer, J., 2002. Clustering and the creation of an innovation-oriented environment for industrial competitiveness: beware of overly optimistic expectations. In Conference paper: International High-Level Seminar on Technological Innovation. United Nations University: Beijing, February.

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

Morrison, A., Rabellotti, R. & Zirulia, L., 2008. Local and global links and the diffusion of knowledge in clusters: a simulation model. In DRUID 25 th Celebration Conference, Copenhagen.

Nigel Gilbert, Andreas Pyka & Petra Ahrweiler, 2001. Innovation Networks - a Simulation Approach. Journal of Artificial Societies and Social Simulation, 4. Available at: http://ideas.repec.org/a/jas/jasssj/2000-25-1.html [Accessed September 9, 2010].

Niosi, J., 2006. Success factors in Canadian academic spin-offs. The Journal of Technology Transfer, 31(4), pp.451–457.

Niosi, J. & Bas, T.G., 2004. Canadian biotechnology policy: designing incentives for a new technology. Environment and Planning C, 22(2), pp.233–248.

Niosi, J. & Bas, T.G., 2001. The competencies of regions–Canada's clusters in biotechnology. Small Business Economics, 17(1), pp.31–42.

Niosi, Jorge & Banik, M., 2005. The evolution and performance of biotechnology regional systems of innovation. Cambridge Journal of Economics, 29(3), pp.343 - 357.

Nohria, N. & Gulati, R., 1996. Is Slack Good or Bad for Innovation? The Academy of Management Journal, 39(5), pp.1245-1264.

Pittaway, L. et al., 2004. Networking and innovation: a systematic review of the evidence. International Journal of Management Reviews, 5-6(3-4), pp.137-168.

Plank, R.E. & Newell, S.J., 2007. The effect of social conflict on relationship loyalty in business markets. Industrial Marketing Management, 36(1), pp.59-67.

Porter, M.E., 1998. Clusters and the new economics of competition. Harvard business review, 76(6), pp.77–90.

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, 41(1), pp.116-145.

Prevezer, M., 1997. The dynamics of industrial clustering in biotechnology. Small Business Economics, 9(3), pp.255–271.

Pyka, A. & Fagiolo, G., 2005. The Elgar Companion to Neo-Schumpeterian Economics, chapter Agent-Based Modelling: A Methodology for Neo-Schumpeterian Economics, Edward Elgar, Cheltenham.

Pyka, A., Gilbert, N. & Ahrweiler, P., 2009. Agent-Based Modelling of Innovation Networks–The Fairytale of Spillover. Innovation Networks, pp.101–126.

Queenton, J. & Niosi, J., 2003. Bioscientists and biotechnology: A Canadian study. In Annual Meeting of the Innovation Systems Research Network. Ottawa, May. pp. 1–2.

Rampersad, G., Quester, P. & Troshani, I., 2010. Managing innovation networks: Exploratory evidence from ICT, biotechnology and nanotechnology networks. Industrial Marketing Management, 39(5), pp.793-805.

Robertson, P.L. & Langlois, R.N., 1995. Innovation, networks, and vertical integration. Research Policy, 24(4), pp.543-562.

Rogers, E.M., 1995. Diffusion of innovations, Simon and Schuster.

Rogers, M., 2004. Networks, Firm Size and Innovation. Small Business Economics, 22(2), pp.141-153.

Rossetti, M.D., 2010. Simulation modeling and arena, John Wiley.

Rothwell, R., 1992. Successful industrial innovation: critical factors for the 1990s. R&D Management, 22(3), pp.221–240.

Schartinger, D., Schibany, A. & Gassler, H., 2001. Interactive relations between universities and firms: empirical evidence for Austria. The Journal of Technology Transfer, 26(3), pp.255–268.

Schiffauerova, A. & Beaudry, C., 2008. Interaction between geographical and technological spaces of collaboration: The Gatekeepers of Canadian biotechnology clusters.

Schiffauerova, Andrea, 2009. Knowledge flows in clusters and innovation networks: The case of canadian biotechnology and nanotechnology. Montreal (Canada): Ecole Polytechnique. Available at: http://search.proquest.com/docview/305138357?accountid=10246.

Schiffauerova, Andrea & Beaudry, Catherine, 2011. Star scientists and their positions in the Canadian biotechnology network. Economics of Innovation and New Technology, 20(4), pp.343-366.

Schumpeter, J. & Backhaus, U., 2003. The theory of economic development. Joseph Alois Schumpeter, pp.61–116.

Van Segbroeck, S. et al., 2009. The coevolution of loyalty and cooperation. In Evolutionary Computation, 2009. CEC'09. IEEE Congress on. pp. 500–505.

Shih, H.-Y. & Chang, P.-L., 2009. Industrial innovation networks in Taiwan and China: A comparative analysis. Technology in Society, 31(2), pp.176-186.

Shmaefsky, B.R., 2006. Biotechnology 101, Greenwood Pub Group.

Smith, D.J., 2007. The politics of innovation: Why innovations need a godfather. Technovation, 27(3), pp.95-104.

Sosa, R. & Gero, J.S., 2005. A Computational Study of Creativity in Design: The Role of Society. AI EDAM, 19(04), pp.229-244.

Souitaris, V., 2001. External communication determinants of innovation in the context of a newly industrialised country: a comparison of objective and perceptual results from Greece. Technovation, 21(1), pp.25–34.

Souitaris, V., 1999. Research on the determinants of technological innovation. A contingency approach. International journal of innovation management, 3, pp.287–306.

Souitaris, V., 2002. Technological trajectories as moderators of firm-level determinants of innovation. Research policy, 31(6), pp.877–898.

Swann, P. & Prevezer, Martha, 1996. A comparison of the dynamics of industrial clustering in computing and biotechnology. Research Policy, 25(7), pp.1139-1157.

Tallman, S., Jenkins, M., Henry, N., & Pinch, S. (2004). Knowledge, clusters, and competitive advantage. The Academy of Management Review, 258–271.

Tian, G. & Zhang, Y.-A., 2008. Simulating the Evolution Process of Industrial Clusters Innovation Networks. In Computer Science and Software Engineering, International Conference on. Los Alamitos, CA, USA: IEEE Computer Society, pp. 532-535.

Udell, G.G., Bottin, R. & Glass, D.D., 1993. The Wal-Mart innovation network: An experiment in stimulating American innovation. Journal of Product Innovation Management, 10(1), pp.23-34.

Velten, K., 2009. Mathematical modeling and simulation: introduction for scientists and engineers, Wiley-VCH.

Wang, N. et al., 2008. Innovations Diffusion on Heterogeneous Newman-Watts Small-World Network: A Computer Simulation. In International Conference on Information Management, Innovation Management and Industrial Engineering. Los Alamitos, CA, USA: IEEE Computer Society, pp. 296-299.

Wu, C. & Zeng, D., 2009. Knowledge Transfer Optimization Simulation for Innovation Networks. Information Technology Journal, 8(4), pp.589–594.

Wu, Chuanrong & Zeng, Deming, 2009. Optimization on Realization Time of Knowledge Demand in Innovation Networks. In Electronic Commerce and

Security, International Symposium. Los Alamitos, CA, USA: IEEE Computer Society, pp. 552-555.

Zaheer, A. & Bell, G.G., 2005. Benefiting from network position: firm capabilities, structural holes, and performance. Strategic Management Journal, 26(9), pp.809-825.

Zander, U. & Kogut, B., 1995. Knowledge and the speed of the transfer and imitation of organizational capabilities: An empirical test. Organization science, pp.76–92.

Zeng, Deming & Wu, Chuanrong, 2009. Decision Support Model and Simulation for Knowledge Transfer in Innovation Networks. In Business Intelligence and Financial Engineering, International Conference on. Los Alamitos, CA, USA: IEEE Computer Society, pp. 519-523.

Zucker, L. G & Darby, M. R, 1996. Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. Proceedings of the National Academy of Sciences of the United States of America, 93(23), p.12709.

Zucker, Lynne G., Darby, Michael R. & Armstrong, J., 1998. GEOGRAPHICALLY LOCALIZED KNOWLEDGE: SPILLOVERS OR MARKETS? Economic Inquiry, 36(1), pp.65-86.

Zucker, Lynne G. & Darby, Michael R., 1995. Virtuous Circles of Productivity: Star Bioscientists and the Institutional Transformation of Industry. SSRN eLibrary. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=225406 [Accessed November 4, 2010].