The Influences and Interactions between Various Scientific Research and Technological Domains in Case of Canadian Nanotechnology

Hadi Shahidi Nejad

Concordia Institution for Information Systems Engineering

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

October 2015

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Hadi Shahidi Nejad

Entitled: The influences and interactions between various scientific research and technological doamins in case of canadian nanotechnology

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

Master of Applied Science in Quality Systems Engineering

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

Signed by the final examining committee:

DR. J. Clark Chair

DR. J. Y. Yu. Examiner

DR. M. Kazemi Zanjani Examiner

Dr. A. Schiffauerova Supervisor

Approved by ______________________________
Chair of Department or Graduate Program Director

____
Dean of Faculty

Date ______________________________

Abstract

The Influences and Interactions Between various Scientific Research and Technological Domains in case of Canadian Nanotechnology

Hadi Shahidi Nejad

In today's world, relationship between domains of science and technology is getting stronger as science contributes to technology in different ways. The interaction between scientific and technological domains is happening in complex innovation processes in which new technological ideas emerge as a result of new discoveries in science. This research aims to investigate interactions between various emerging scientific and technological domains and their influences in the development of both patents and publications in the field of nanotechnology.

The study uses real data of the journal articles and patents in nanotechnology between 1995 and 2008 which we clustered into scientific and technological domains. In clustering phase, terms and phrases were located in each record using singular value decomposition algorithm and then documents were assigned to cluster labels by applying standard vector space model algorithm on them. To achieve our research goals, in next step, we built the network of nanotechnology article-patent citations and investigated various network topological parameters over all nodes. The patent-article network is built on citation links among different nodes of patents and articles, while patent nodes cite a set of NPLs (Non-Patent Literature) and NPLs are also citing another set of articles as their references. Focusing on the role of NPLs, we studied trend of network topological parameters like betweenness centrality and degree centrality while looking at correlation between them. We highlighted leading patents in technology and leading articles in NPLs and their cited articles set which could be seeds of innovation in nanotechnology.

Our main results of this research are focused on the role of NPLs as gate-keeper nodes in bridging ideas from scientific to technological domains. Comparing NPL citation counts to articles and patents, results show higher range of NPLs' contribution to the development of scientific fields than technological domains in Canadian nanotechnology. We also highlighted most cited and citing NPLs nodes of article-patent citation network as significant nodes in connecting science and

technology. Using average of citations per article metric, we calculated the rate in which different technological domains influenced by scientific NPLs and also the impact of NPLs on development of different scientific and technological domains. Regarding the contribution of top cited NPL articles in development of scientific domains, we discovered a positive correlation between citation count and betweenness centrality measure of articles, which indicates the more an article is cited by patents and other articles, the more influence it has on the transfer of ideas from scientific to technological domains. We also observed that citation count value of journals in our citation network has a positive relation with the number of scientific domains it contributes to. In addition, we discovered the positive relation between patent citations count and journal's impact factor. This is interesting to us since we can see the more articles of a specific journals are cited by patents, the more impact factor the articles of that journal have. In other words, impact factor not only shows the impact of articles on development of scientific domains, it also shows how the articles of a journal have impact on development of technological domains. Regarding the NPL's contribution to development of technological domains, we found a positive relation between NPL journals' citations to technological clusters and the number of technological clusters they cover. Results showed the increasing trend of journals' contribution to different technological domains as citation count value of journals increases.

It is worth to mention that this study is the first to examine the flow of ideas from scientific to technological fields which uses a citation network of both patent and article nodes, and investigating leading articles and patents which play a crucial role in keeping this knowledge flow alive in nanotechnology related sub-fields.

Acknowledgements

I take this opportunity to express gratitude to all who provided me with all the necessary facilities for the research. I would like to express my sincere thanks to my supervisor, Dr. Schiffauerova. I am extremely thankful and indebted to her for sharing expertise, and sincere and valuable guidance and encouragement extended to me.

I also place on record, my sense of gratitude to my parents and my brother for their endless support and encouragement. Finally I would like to express my sincere thanks to my colleagues in our research team and friends for being with me throughout this program.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction and Literature Review

Introduction

Over the past decade, network analysis and study of network parameters and features has been one of the most interesting scientific fields of research. Network analysis is being used extensively in a wide range of applications and disciplines such as mathematics, physics, chemistry and sociology. Among all these applications, recently scholars are more interested in the study of innovation processes which also shows the importance of network analysis methods in today's research.

Social networks help us to understand how ideas, innovations and patents are interacting with each other. Research in this field can investigate how individuals contribute to their social context and how their relations are being influenced while collaborating with others. Hence, social networks can help in development of innovation processes and increase opportunities for learning Kolleck, 2013).

Social networks can be used in investigating change processes in more detail and study knowledge diffusion. Moreover, by the help of network analysis techniques researchers can study existing networks and identify innovation potentials in order to generate new information and see how structural and topological aspects of networks are related to the growth of technological and scientific domains (Kolleck, 2013).

In this thesis we study change and innovation through deep look at network relations between different technological and scientific sectors. For this purpose, the study and analysis of networks will help us to find leading actors in citation network of articles and patents. Regarding the inter-relation links between different types of nodes in a network, we investigated how much a group or an individual node contributed to the growth of scientific and technological fields.

Our research is based on topological network analysis of real data taken from nanotechnology scientific and technological references in Canada. The importance of nanotechnology in Canadian industry and economic is significant, and recently inventors, authors and entrepreneurs have been paying more attention to this field. It is a relatively new sector which is growing fast in Canadian industry. Also, nanotechnology as an innovative science and technology branch has a significant contribution to science advancement and can have a positive influence on economic growth in the world and specifically in Canada.

According to the report of Canadian government website, Canada is one of the leading countries in nanotechnology R&D infrastructure. They achieved this by support of highly skilled resources, growing number of companies involved in nanotechnology, and responsible government committed in development of nanotechnology (Perez, 2013).

This thesis will investigate the influences and interactions between various research and technological domains in the development of Canadian nanotechnology, focusing on innovation networks' structure and knowledge creation in technological and scientific levels. The results of our research can serve as a basis for the design of governmental policies or organizational strategies related to knowledge creation in various nanotechnology scientific and technological domains. The main aim of this thesis is to present a better view of citation knowledge networks in Canadian nanotechnology which can improve the understating of interactions between various technological and scientific fields in this sector.

Literature Review

In this section, first we will discuss articles focused on linkage between science and technology. In the next section we will take a closer look into articles which used citation counts as measure of articles and patents significance. These articles mostly did citation analysis over networks of patents to track the flow of innovative ideas and map science into technology. We follow this chapter by discussing research works focused on nanoscience and nanotechnology articles and patents. These research studies used different methodologies over data-sets patent and article records to cluster scientific and technological domains in nanoscience and nanotechnology. In the fourth section we discuss research papers related to nanotechnology in Canada. Finally we will discuss research gaps in regards to literature review and chapter summary will come at the end.

1.1 Science and technology Interactions

The relationship between science and technology has been investigated for a long time. Back in 1965, De Solla Price (1965), developed a two-stream model based on citation analysis of science and technology journals. This model focuses on the autonomy of science and technology as cognitive systems. By tracing citations in science and technology journals, De Solla Price (1965) studied separate cumulative structures with scientific knowledge building on old science and technology on old technology. He also detected a weak and reciprocal interaction between science and technology and in addition, he found the closest interaction between science and technology taking place in the period of education when 'budding scientists read the archival literature in their fields' (De Solla Price, 1965).

Linkage between science and technology is investigated in wide range of research works as scientists used variety of methods in tracing knowledge into technology. Carpenter et al. (1980), studied patent-to-paper citations in 319 gas laser patents and 399 prostaglandin patents from the USPTO[1] and found that nearly 90% of all journal references made by patent applicants and examiners refer to basic or applied scientific journals, as opposed to engineering and technological literature. They also found that the time between publication of a journal article and the patent application citing that article was relatively short like three to five years. In addition, patent applicants and examiners tend to cite scientific articles in the central core of the scientific literature covered by the Science Citation Index (SCI).

An interesting study on interactions between science and technology continued as the theory of 'pipeline' model introduced by Brooks (1994). According to this model, the innovation process in which new technological ideas emerge is a result of new discoveries in science and move through a progression from applied research, design, manufacturing and, finally, commercialization and marketing (Brooks, 1994).

Following up research studies in the same scope, Nelson (1992) defines innovation as the processes by which firms do product designs that are new to them, whether or not they are new to the universe, or even to the nation, while the current models of innovation often emphasize on originality in the sense of newness to the universe. Yet research and development is also necessary

[1] US Patent Office

for learning about technology even when it is not ‘new to the universe’ but only in the particular context in which it is being used for the first time (Brooks, 1991).

The relationships between science and technology are very complex, and interactively different in different fields and at different phases of a technological ‘life cycle’ (Brooks, 1994). According to Nelson (1992), technology can be defined both as “...specific designs and practices” and as “generic knowledge... - that provides understanding of how [and why] things work...” while he defines innovation as “...the processes by which firms master and get into practice product designs that are new to them, whether or not they are new to the universe, or even to the nation.”.

The linkage of science and technology fields was also investigated by Narin and his colleagues (1997), as they focused on citation count of non-patent references as one of the measures of citation analysis. This research was interesting to us as it is based on analysis of 430,226 non-patent references which were listed as 'other references cited' on the front pages of 397,660 U.S. patents issued in 1987-1988, and 1993-1994. They investigated the linkage between science and technology fields by looking at different categories of non-patent references while looking at patent and article links in a specific section. Finally results of this research shows public science plays an essential role in supporting U.S. industry, across all the science-linked areas of industry, amongst companies large and small, and is a fundamental pillar of the advance of U.S. technology (Narin, et. al., 1997).
In order to understand the complexity of the relationship between science and technology, it is first important to establish how the two domains differ. According to Compton (2004), the differentiation between science and technology can be characterized by three key factors:

- The domain’s core business (its purpose)
- Its view of what ‘exists’ in the world (its ontological stance)
- How it defines and validates knowledge (its epistemology)

To have a better idea of this type of interactions first we are going to make a brief overview of science and technology.

- **Science**: The main purpose of science is to explain the natural world through investigative practices that involve observations and controlled manipulations of that world. Science can be most comfortably argued today in terms of a ‘critical realist’ stance. This reflects a view that ‘things’ exist in the world and ‘are as they are’ (Lopez and Potter, 2001).

- **Technology:** According to Compton (2004), the purpose of technology is to intervene in the world to produce something ‘other’ to that which currently exists. It achieves this through iterative intellectual and design-based practices that involve multiple sources of input. These input sources include a mixture of that defined as natural, material, simulated and conceptual.

As discussed so far, science and technology interactions are happening in different complex ways which depend on nature of scientific and technological domains, while there are many parameters which have effect on it. But to have a quick summary, here we mention Brooks’ (1994) findings on six ways in which science contributes to technology:

1) New knowledge which can directly flow into technological patents as their main source of idea
2) As a source of tool and technique in engineering efficient designs
3) Analytical methods and laboratory techniques in scientific methods which can lead to design and invention of patents
4) Use of research as a source of absorption and development of human skills in technology
5) Creation of knowledge base which is important in evaluation of technology and it’s social impacts
6) Creation of knowledge base which helps more strategies of applied research in development of new technologies Brooks (1994, P 447)

According to Chen et. al. (2002) tracing the transfer of knowledge from science to technology, from technology to technology, is the most popular area of research in relation to citation analysis.

In one of the other research papers of Chen (2004) on the topic of linkage between science and technology, he mentions both scientific publications’ and patents’ citations as the most fundamental indicators of impact. However, tracing knowledge diffusion between science and technology domains is still a challenging issue. Conventional approaches are often qualitative in nature, including interviews, questionnaires, and in-depth case studies. Such methods are often time-consuming, expensive to use, and requiring a substantial level of prior domain knowledge.

Also Chen and Hicks (2004) illustrated a useful approach, combining statistical mechanics of complex networks, network visualization, and citation analysis. The goal of their research was to improve the understanding of knowledge diffusion and technology transfer, especially with principles and streamlined methodologies for citation analysis, and the expanded scope of citation analysis. The interesting point about their research was to discussion on some of the issues concerning knowledge diffusion and how to trace the process of knowledge diffusion by utilizing patent citation networks.

There are often multiple factors that may influence the path and direction of knowledge transfer between specific scientific and technological sectors. In fields such as health and semiconductor research there tends to be a strong positive connection between basic research and technological innovations, whereas in fields such as information technology it is technology that leads science by more than a year according to the publication dates of cited patents and scientific publications (Hicks, et. al., 2001).

One of the most recent interesting research papers on tracing links between science and technology was done by Karvonen and Kassi (2013). Their research discuses impact of science on technology application in converging technological environments by analysis of 464,225 patent application and 506,225 NPLs between 1978 and 2006 in paper industry. This research was interesting to us as they used the idea of separating NPL citations into “scientific” and “technological” groups and measured the influence of science of development of technological patents in their scope. The results of their analysis reveal great differences in the “science intensity” between different industry sectors. The patent indicators and the detailed analysis of NPL citations give an insight into interaction between technological and science convergence.

Articles on science and technology linkage which we discussed so far, used variety of data sets and ideas to study the way these types of interactions can happen. In the next section, we focused more on research works that used article and patent data to study the science and technology mappings.

1.2 Citation Networks of Innovation and Science Mapping

In this section we are going to start by some definitions of basic terms of our research such as innovation and complex networks. Then we will continue by discussing researched focused on citation network as their method of science and technology mapping.

Innovation is a new idea, device or process. It can be viewed as the application of better solutions, in achieving new requirements of existing market needs. Innovation can be spread from the innovator to other individuals and groups. Researchers study this type of relations by building and analysis of networks of innovation to have a deeper look at how this knowledge diffusion process takes place among technological and scientific fields, inventors, assignees, authors and geographical regions (Marienville, 1992).

A network is composed of nodes which are related to each other by links or edges. Depending on the type of network, edges can be directed or undirected. Based on the research by Wasserman and Faust (1995), by extracting patterns in network relation between nodes like inventors, authors or firms we can study how these actors are interacting with each other.

Studies in citation analysis field go back to 1973, when Henry Small (1973), published his work on co-citation analysis which became a self-organizing classification system that led to document clustering experiments and eventually what is called "Research Reviews" (Kas. M., 2011).

Egghe and Rousseau (1990) explain a citation network as a directed graph which is consisted of nodes and directed edges. According to their definition, when a document di cites a document dj we ca n show this by an arrow going from the node representing di to the document representing dj. This collection of arrows and nodes builds the citation network.

Tijssen and Van Raan (1994), presents basic principles and examples of representations derived from the analysis of co-occurrence frequency data to bibliographic information elements, such as key words and citations, in research publications and patents. According to their research, these bibliometric maps provide a means for communicating information on relational features of the science and technology either for analytical or representational purposes. As the final result three empirical examples of science maps were presented with a focus on their application for impact assessment in both scientific as well as technological fields focusing on collaboration within Dutch research on coal and coal products. This research was interesting to us in case of creating both science and technology maps and investigating their interactions within a specific field of industry.

Lawrence et al. (1998) introduced autonomous citation indexing, which enabled automated extraction and grouping of citations for academic/scientific documents. This was a great step through automatic use of citation measures in online databases today, as citation extraction was a manual process before that.

In a National Academy of Sciences colloquium entitled "Mapping Knowledge Domains" (Shiffrin and Börner, 2004), the term "mapping knowledge domains" was used to "describe a newly interdisciplinary domain of science which can be used in charting, mining, analyzing, sorting, enabling navigation of, and displaying knowledge". Citation networks can also help researchers identify relations between topics of different subfields and topics by looking at networks structures around specific topics. According to Shiffrin and Börner (2004), the value of mapping knowledge domains extends beyond the bounds of information science, to scientists, researchers, governmental institutions, industry, and members of society. As authors also emphasize, although the extraction and organization of knowledge may form the scientific core of scientific fields, the results will not be useful unless the user can understand and interact with the mapping systems. Knowledge typically is organized along many dimensions, but a map with thousands of dimensions cannot be used effectively.

We also can study the evolution of technological and scientific fields over time by studying changes in semantics and structure of citation networks. According to Rocco and his colleagues (2009), "knowledge mapping" or "science mapping," based on citation network analysis and information visualization, has become an active area of research that helps reveal such an inter-connected, network of scholars and their seminal publications and ideas. Also according to Chaomei Chen in his book, Mapping Scientific Frontiers (Chen, 2003), science mapping helps to track the relations between research fronts, which are leading areas of research. Such maps can also simply be used as a method to track the way in which research areas are distributed. According to Rocco and his colleagues, by using a series of sequential maps, we can see how knowledge advances. Mapping scientific frontiers involves several disciplines, from the philosophy and sociology of science, to information science, science metrics, and information visualization (Roco et.al. 2009).

Jaffe and Trajtenberg (2002), studied the flows of knowledge by modeling the flow of patent citations over time and across institutional and geographical boundaries. They used data-set of 88,257 patents granted between 1963 and 1990 and assigned to United States corporations and found Within-country citations are more numerous and come more quickly than those that cross country boundaries.

Meyer (2001) published a research paper on patent citation analysis of nanotechnology and nano-science. He investigated interrelationships between science and technology in the field of nanoscience and nanotechnology by tracking patent citation relations at the sectoral disciplinary, the organizational, and the combined industrial/organizational levels. He also investigated the geographic location and organizational affiliation of inventor/authors. According to his results, there are only a small number of citations connecting nano-patents with nanoscience articles, while nanoscience and nanotechnology appear to be relatively well connected in comparison with other fields. He observed that, nanoscience and technology are still mostly separated spheres, even though there are overlaps. As an analysis of title words shows, he also found that university-assigned patents seem to cite papers more frequently than other patents.

Al-Thubaity and Ahmad (2002), studied domain of nano-structured tunnel diodes in semiconductor physics based on patent descriptions retrieved from USPTO. They examined sets of terms in order to identify the patterns in them and to understand how knowledge evolves in an emergent domain. However, much of their work was carried out manually, which tends to be limited in terms of flexibility, cost-effectiveness, and scalability.

Interesting point about Meyer's research was his work on nanoscience articles using citation analysis methods. But considering his data-set, he used 5,400 nano-articles published between 1991 and 1996 which is covering just a period of 7 years in nanoscience articles. Also the number of articles in data-set is relatively small.

Some researchers such as Jaffe and Trajtenberg (2002), used patent citations to investigate the diffusion of technological information across institutions over time. Such works involve econometric analysis, and parameter estimation and can help investigate knowledge diffusion. They developed a model of the process generating subsequent citations to patents to have a better view of knowledge diffusion. As we also mentioned before in our literature review, their results indicate that diffusion is geographically localized. Controlling for other factors, within-country citations are more numerous and come more quickly than those that cross country boundaries. According to Jaffe and Trajtenberg (2004), citation patterns across technological fields conform to prior beliefs about the pace of innovation and the significance of "gestation" lags in different areas, with Electronics, Optics, and Nuclear Technology showing very high early citation but rapid obsolescence, whereas Drugs and Medical Technology generate significant citations for a very long time (Jaffe and Trajtenberg, 2004).

According to the literature of knowledge mapping it is clear that domain visualizations and the ability to interact with knowledge and view it from a variety of perspectives play a critical role in having a better understanding of knowledge diffusion. The results of different algorithms used to extract and organize relevant data in research can be displayed in many ways. For example, maps might follow major researchers, most cited articles and books or most cited patents to point to emerging trends or articles or patents organized into scientific or technological domains over time. In the next section, we will discuss research studies which focused on nanotechnology publications and patents as their data-sets.

1.3 Nanotechnology Publications and Patents

One of the definitions of nanotechnology is as the understanding and control of matter at the nanoscale (from approximately 1 nanometer to 100 nanometers in length). It includes nanotechnology science, engineering and technology, and also imaging, measuring, modeling and manipulating matter at nanometer scale (Perez, 2003).

Haung and his colleagues (2002), have a research on bibliographic analysis of references in patents. The main target of this research was to investigate the linkage of science into technology in nanotechnology. They used data-set of 1115 EPO and 514 USPTO nano related patents and found a steep increase in the number of patents and non-patent references (applications and grants) during the early beginning of the 1990s. According to their results the patenting performance in nanotechnology has been contrasted with the publication performance in the science base. In quite some science domains we see that EU-15[2] perform very well, in several instances even better than

2 EU-15: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain, Sweden and the United Kingdom.

the US and/or Japan. This research article was interesting to us because of the methodology that authors used in analysis of citations while considering non-patent references as one of the analysis measures despite the low number of patent records was one of the limitations of this study.

Analysts have stated that nanotechnology will lead to the next industrial revolution by development in a new era of manufacturing and engineering. According to O'Brien and Cummins, (2011) nanotechnology is completing a twenty years transition from lab to market, similar to a historical pattern previously seen in fields such as plastic materials and biotechnology and more than 50 billion USD in products sold worldwide in 2006 incorporated nanotechnology with very diverse applications. National Science Foundation (2001), reported that the projected worldwide market size of nanotechnology will top $1 trillion USD annually by 2015. Consequently, social scientists devoted a great deal of energy to studying the characteristics of emerging technology and its economic and societal implications. As a result of its great potential, nanotechnology has become the focus of science and technology policy in various countries and transnational organizations (Haun and Hammer, 2008).

Looking at nanotechnology related research papers, back in 2002, the European Commission sponsored the scholars, residing at Leiden University in the Netherlands and Fraunhofer ISI in Germany to employ a more robust methodology to identify centers of excellence in Europe in the field of nanotechnology (Noyons et al., 2003). In these studies, a set of bibliometric indicators that address the inter-disciplinarily of nanoscience and nanotechnology were developed to assess the performance of researchers and institutions in Europe. Using indicators such as the average number of citations per publication normalized by traditional science areas, the authors were able to correct the bias of the evaluation which resulted from higher probability of being cited in the basic sciences than in the applied sciences (Haun and Hammer, 2008).

Focusing on the United States and the use of USPTO patent data, Huang, Chen, Roco and co-authors analyzed the general trends of nanotechnology research and development, the leading players, and the evolution of technology topics with respect to countries and institutions in the field (Huang et al. 2004).

There is a large number of empirical research works which used citation networks in study of knowledge and technology evolution. Bassecoulard, Lelu, and Zitt (2007) used the methodology of citation analysis to obtain a database of all the nanotechnology publications from 1999 to 2003. They also used cluster analysis to classify the literature into different domains according to the similarity of the papers in the references which is, the source of information. This is one of the main references of this research in terms of methodology as they did nanoscience mappings based on citation flows. According to their results, fifty themes have been extracted and further aggregated in seven super-themes as higher level clusters. As authors mentioned, the main limitation of their work was the exploration of the map and themes performed only with a basic characterization. Also the data set which is used by Bassecoulard, Lelu, and Zitt (2007) was limited to a five year timeline period which is not covering as many years as similar research studies in this scope. Igami and Saka (2007) carried out a citation analysis, mapped the nanotechnology field and classified the nanotechnology publications into 30 subfields. Their analysis showed a multi-disciplinary character of some fields such as ‘nano materials and devices’, ‘genomics’ and ‘environment’. Their research included a geographical analysis of different countries share of research in nanotechnology scientific fields. This research was interesting to us in terms of finding how knowledge is evolving not only across disciplines but also across countries and regions. To map the world’s nanotechnology scientific publications for the years between years 2002 to 2006, Leydesdorff and Wagner (2013) focused on the ten core journals in the field. They demonstrated that more than one percent of its world share of nanotechnology publications per year is decreasing. One of the interesting articles which investigated the linkage between science and technology using citation count measure and citation analysis, was done by Szu-Chila Lo (2007) as patent count and the research linkage is examined by tracing the non-patent citations. 1,048 USPTO patents granted to Japan, Korea and Taiwan from 1976 to 2004 in genetic engineering including gene mutation, cell fusion, genetic modification and recombinant DNA and 2,006 referenced patents cited by those 1,048 patents were examined in this study. The author further constructed the linkage foundation between public science and technology development by examining 10,230 non-patent citations. Results of this research show that the public science has high impact on the technology development of Genetic Engineering Research. Based on the number of citations, the results imply that the public science provides foundation for the Genetic Engineering Research and strong linkage between the private and public sectors. Also the scholarly journals are the valuable sources among various communication channels for research output for technology development. The titles highly cited in the patents have similar significant influence on the public science since they were also heavily cited by other scholarly journals. It is worth to have further studies on the literature level to reveal the linkage among public institutions and private organizations (Lo, 2007).

The research work by Lo (2006), was one of the most inserting publications we reviewed in terms data and methodology. Similar to the main objectives of our research, this study used both patent and non-patent citation counts of USPTO patents as measures of citation analysis. Lo also investigated contribution of leading scientific journals in genetic engineering and results show Most of the highly cited titles cover the papers with the topics in Biochemistry and Molecular Biology, Genetics Research. Nature and Science that are multidisciplinary titles were also heavily cited.

In another research paper by Lo and Chiu (2009), he analyzed 213 USPTO patents granted during the period of 1985 to 2008, which were identified as nanotechnology patents by IPC Numbers. The 4,161 cited patents and 4,593 cited non-patent literatures referenced by 213 nanotechnology patents were included in the study. The study took bibliometric approach as Patent Count was used to show the research productivity and Citation Count was applied to reveal the linkage of science research and technology development. The results of this research showed with local advantage, the United States is the leading country both in productivity and research impact in nanotechnology research, With Northwestern University as technology base, nano-sphere has leading position in Molecular nanotechnology. Research outcomes of both public science and industrial technology play import roles in the development of nanotechnology and finally scientific linkage could use as an indicator to show the value of the research works in public science (Lo and Chiu, 2009).

1.4 Nanotechnology in Canada

Having a quick look at technology time line, we can find the fifth technology revolution in the world is happening based on nanotechnology and molecular manufacturing (Perez, 2003).

In 21st century, nanotechnology patent literature has a considerable growth. The early 2000s correspond to a period during which several government initiatives worldwide increased nanotechnology research and education funding (Jordan, et. al, 2014). In this research also we analyzed the growth of both patent and article literature in Canadian nanotechnology in a time period between 1995 and 2008. Our results also confirm that by the first decade of 21st century we see more activities from inventors and authors in this field.

According to 2013 patent literature review by Jordan et. al. (2014), the total number of publications in nanotechnology patent literature increased 5 percent in 2013 and has more than tripled since 2003. Being more specific, by taking a closer look at the innovation trends in graphitic carbon-based nanotechnology –which in this research is one of our main scientific and technological clusters -, this field has unique structures that have applications in wide variety of other fields like electrical, spectral, thermal and mechanical properties and that is why recently we have more inventors and authors interested in this nanotechnology field. Regarding the share of Canadian inventors in nanotechnology patents, according to Dang, et. al. (2009), in Figure 1, we can see high contribution of Canada to the number of patent publications especially after year 2003.

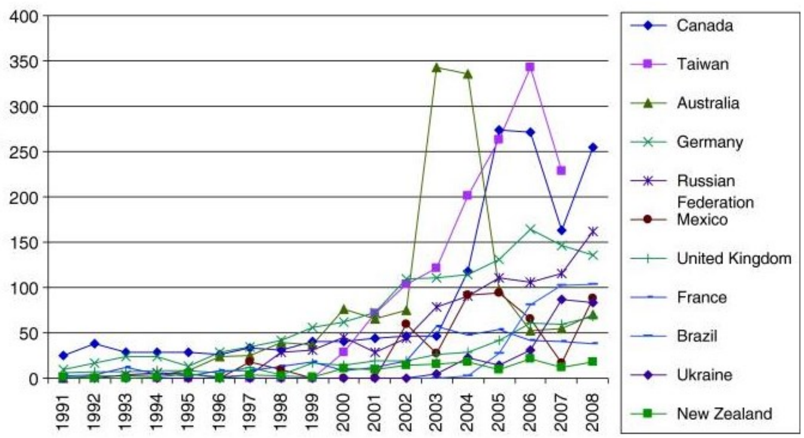


Figure 1: Canadian nanotechnology patent applications comparing to 11 patent office's using "title–abstract" search from 1991 to 2008 (Dang, et. al., 2010)

As nanotechnology innovation grows, several regions have emerged as leaders in this important area of technology. In this thesis also we analyzed the location of leading regions in Canadian nanotechnology patent and article literature and how they interact with other regions on case of technological and scientific collaborations.

Based on Canadian government website (2012), recently the number of intergovernmental and other international organizations concerned with the responsible development and application of nanotechnology is increasing in Canada. This is how other countries also are motivated to have shared interest in nanotechnology.

Nanotechnologies offer the potential for innovation development and leading in industrial competitiveness which leads to economic growth and social benefits in Canada (Jordan, et. al. 2013).

According to Canadian government official website report (2012), nowadays, government of Canada realizes the importance of nanotechnology and hence invests great amounts in this technology. Based on Canadian government official website report, in 2005, ISO and IEC, the two major international standards development organizations launched initiatives to facilitate the safe and responsible development and use of nanotechnologies. There are also two technical committees established with some 45 member countries, including Canada, and 38 liaison groups. It is worth to mention that, since 2005, Canada has held the leading international convener roles for the Terminology and Nomenclature groups under both committees

Research activities play a crucial role in development of nanotechnology as the development of standards is taking place concurrently with research and commercialization activities rather than product development as in more established fields.

One of the interesting research papers in this domain as it focused on Canadian nanotechnology patents and building innovation network of Canadian patents based on technological clusters is done by Schiffauerova and Beaudry (2009) where they studied innovation in Canadian nanotechnology clusters and networks on a data from the intersection of the Nano-bank database for Canadian inventors. In this research 8 Canadian nanotechnology clusters were identified and then the Canadian nanotechnology innovation network was built to describe the collaborative behavior of the inventors. According to this article results most collaborative activity is related to inside nanotechnology clusters and Canadian inventors who decide to build cooperation outside their clusters most often prefer to do so with collaborators from abroad, mainly from the USA.

Schiffauerova and Beaudry (2011) published another research paper focusing on the impact of collaboration and co-inventorship network characteristics of Canadian nanotechnology inventors on the quality of their inventions. They investigated the impact of four types of variables on patent quality and showed that the presence of more central inventors and of star scientists in the research team has a positive influence on patent quality, while repeated collaboration has a negative impact.

In the next section we will bring a summary of most related articles to our research in terms of authors, year, data-set, research target and results. Afterwards we will identify research gaps based on the literature review.

Table 1: Most related articles of literature review

Authors	Methodology	Year	Data	Research Target	Results
Tijssen Van Raan	Bibliometric Co-Occurrence Analysis	1994	- 29,019 research publications over the period 1991-1992	Relational features of the science and technology	Three empirical examples of maps presented with a focus on their application for impact assessment in both scientific as well as technological fields
Narin Hamilton Olivastron	Citation Analysis	1997	430,226 non-patent references which were listed as 'other references cited' on the front pages of the 397,660 U.S. patents	Examination of the contribution of public science to industrial technology	Public science plays an essential role in supporting U.S. industry
Meyer	Patent citation analysis	2001	5,400 Nano Articles published between 1991 and 1996	Interrelationships between science and technology in the emerging area of nanoscience and technology	Nanoscience and technology are still mostly separated spheres
Jaffe Trajtenberg Fogarty	Statistical Analysis	2002	USPTO Patents (1985-1993)	Importance of their inventions, the extent of their communication with other inventors	Significant correlation between the number of citations a patent received and its importance
Verbeek Callaert	Statistical Analysis	2002	1115 EPO and 514 USPTO Nano-related patents	Investigate the presence of scientific research in the “prior art” description of a patented invention	US holds the highest number of Nano-patents, followed by EU-15 and the Developed Asian countries
Jaffe Trajtenberg	Statistical Analysis of Citation info	2004	88,257 patents granted between 1963 and 1990 and assigned to United States corporations	Flow of Patent Citations over time (Geographical)	Within-country citations are more numerous and come more quickly than those that cross country boundaries.
Chen Hicks	Citation Analysis	2004	267 U.S. patents which in turn made 5,387 patent-to-science citations to	Improve the understanding of knowledge diffusion and technology transfer	Illustration of a useful approach by combining statistical mechanics of complex networks, network

			562 unique scientific articles		visualization, and citation analysis
Bassecoulard Lelu Zitt	Citation analysis	2007	All nanotechnology publications between 1999 and 2003	Citation Based Mapping of Nano sciences field	Extraction of different Nano science sub field using axial k-means clustering method
Leydesdorff	Social Network Analysis	2007	7,379 journals included in the Journal Citation Reports of the Science Citation Index and the Social Sciences Citation Index 2004	Topological network Parameter values impact on Interdisciplinary of journals	The finding of betweenness centrality as a possible indicator of interdisciplinary in journal mapping.
Miller Fern Cardinal	Statistical Analysis	2007	211,636 patents from 1,644 companies during the period 1985–96	Use of Knowledge in technological innovation within firms	The positive effect of the use of interdivisional knowledge on the impact of an invention is stronger than the effect of using knowledge from outside firm boundaries.
Haung Notten Rasters	Bibliometric Analysis	2008	120 social science studies in nanoscience and technology	Comparative analysis of bibliometric search strategies	Ranking tables of the top ten nanotechnology subject areas, the top ten most prolific countries and institutions
Lo	Bibliographic analysis	2009	1,048 USPTO patents granted to Japan, Korea and Taiwan from 1976 to 2004	Investigate the linkage between Science and technology in generic engineering	Public science does have research impact on the technology development, in particular for Genetic Engineering Research
Karvonen, Kässi	Statistical Analysis on citation data	2013	464,225 patent application and 506,225 non-patent references (NPR) in the period 1978-2006	Impact of Science on Technology Applications in Converging Technological Environments	Found a great differences in the “science intensity” between different industry sectors.

Having a closer look at the summary of our literature review we can see that we have various research works on the interactions between science and technology domains, these research studies used variety of methodologies such as bibliographic analysis, statistical analysis, citation analysis, etc. to analyze their data-sets. As we were focused on citation count of article and patents as a measure of article/patent quality we had a closer look into some research articles such as Meyer's article on 2004 using citation flows to investigate interactions between scientific and technological domains. The methodology of research studies using citations were closer to our objectives as they investigated science and technology considering more similar indicators to discuss science and technology significance.

As summarized in Table 1, we can see most of these research works were focused on engineering domains and some on nanoscience related fields. Most of research targets were focused on geographical analyses of patents citations. The data-sets used in the citation analysis investigations cover time lines of less than 10 years. Low number of records and the fact that they were using either set of patents or set of articles for their investigation leads to some research gaps which will be identified later one.

We this decided to use data-set of both articles and patents between years 1995 and 2008. We used citation count of article and patents as a measure of their significance in citation network and used extracted both science and technology clusters to map the scientific and technological overlaps and track each leading field growth over research timeline. As one of the main objectives of our thesis, we will try to fill in gaps by investigating interactions between scientific and technological domains by highlighting top domains of science and technology in Canadian nanotechnology. We will discuss objectives of our thesis in the next chapter.

1.5 Chapter Summary

In this chapter we summarized the main surveys and research articles which we used as references of choosing the most proper methods of data analysis. We also brought a brief summary of most interesting articles or our literature review in Table 1 discussing their year of publication, data, target of research and results. And finally at the end of this chapter we mentioned the research gaps in our scope of research we found regarding literature review and we are trying to fill in this thesis.

2. Research objectives, Data and Methodology

Introduction

In this chapter we will we will start the discussion on our research objectives, and will continue the chapter discussing research methodology steps along with the flowchart of the whole research process. Regarding the research methodology section, we will start by having a closer look at our data-set and data pre-processing step. We will continue to next sub-sections by discussing our cluster analysis method, patent-article citation network structure and finally network topological measures which will help us to have a better view of analyses done in next chapters.

2.1 Research Objectives

In this section we review the main objectives of our research in this thesis:

- Investigate the influences and interactions between various research and technological domains in the development of Canadian nanotechnology, by detecting the emerging and promising research and technology domains

To reach the main objective, we need to reach these sub-objectives in different steps:

1. Investigating most important (Citing/Cited) NPL[3]s as they play gate-keeper role in connecting scientific and technological domains and study the contribution of NPL articles in development of these fields
2. Study of data cluster maps and investigating growth of scientific and technological clusters in research time line
3. Investigate impact of top journals and Canadian institutions in development of scientific and technological domains
4. Investigate the relationship between topological network parameter values such as in-degree and citations per articles, and journals' impact factor

[3] Non-Patent Literature

5. Examining most citing patents and their technological domains as they play a significant role in keeping technology and science domains related

In the next section, we will discuss each of our research methodology steps in detail.

2.2 Research Methodology

In this section we will discuss the methodology of our research. We will start by discussion about data description and preprocessing step, while having a closer look at NPL data definitions. We will continue the chapter by discussing data clustering phase as we focus on vector space model clustering algorithm. In the next section we will discuss the article-patent network structure and we will explain different types of citation links between network nodes. And finally in network analysis sub-section, the network topological measures are discussed. The main research methodology steps which we will discuss are as below:

- Data preparation and pre-processing
- Data clustering
- Network build
- Network analysis
- Results interpretation

2.2.1 Data Description and Preprocessing

In the following sections, data-set records' description, data preparation and pre-processing steps will discussed.

2.2.1.1 Data Description

As we mentioned in section before, our data set includes records of patents, NPL articles and their cited article records.

Regarding Patents data, we used the core set of 4,522 distinct Canadian nanotechnology patents from 1995 to 2008 including their NPL citation information in our data-set.

Our main data-set of articles includes 579,214 article records with 3,626,281 records of cited articles. A core set of nanotechnology articles collected from Scopus database[4] is matched with "Web of Science" database to extract the standard article information mapped with WOS[5] format. In total, regarding the citation relations between articles, our article citations data table included 16,104,188 records of citations. We used a subset of these article records in building article-patent citation network as only some of them were being cited by NPLs.

Our NPL records, which were used to build the middle tier of the network[6] involve 12,353 NPLs with 9,712 records polished in 1995-2008 timeline. This set of NPLs in also connected to a set of cited papers with 2,621,888 articles published between 1995 and 2008.

We will continue the discussion on data-set focusing on NPL data in more details as follows in the next sub-sections.

2.2.1.2 NPL data preparation

In data preparation, as mentioned before, cleaning and standardization of nanotechnology articles data was done. Focusing on NPL data, these sub steps were needed to be done in data preparation phase:

- Extracting "non-patent literature" or NPLs from each patent citation
- Extract the bibliographic information of each NPL
- Match the bibliographic information with article publication database

We will discuss these steps in detail, while reviewing NPL definitions.

Non patent literature of a patents can defined as those documents and publications that are not patents or published patent applications, but are cited as references for being relevant in a patent invention. For instance, a magazine article or doctoral thesis relevant to a claimed invention can be cited as non-patent literature. In other words, any technical document that is neither a patent nor a patent application and that is submitted by a party or cited by an examiner during patent

---

[4] This data was extracted from SCOPUS database by another research team member using specialized keywords related only to nanotechnology. His research led to thesis: "A Network Perspective of Nanotechnology Innovation: A Comparison of Quebec, Canada and the United States", 2012.

[5] Web of Science

prosecution[7] is a non-patent literature. The non-patent literature includes especially scientific papers used as prior art to show that an invention claimed in a patent or patent application was known or obvious before the filing of the application (Glossary of patent law terms, 2015).

NPL records play an important role in our citation network as they keep nodes of patent and article subnetworks connected. We received our NPL data set as citation information of nanotechnology patents in raw strings format. These strings included NPL information such as title, authors, year, type, publisher, volume, start page, end page and etc. To parse our raw NPL string, first we needed to decompose each string into different fields. Regarding the citation template on Freecite library, we parsed all raw NPL strings into clean and separate fields (Freecite, 2009).

Here is one sample string and parsed strings mapped on parsing rule:

Example: S Ghosh, AK Sood, N Kumar (2003), Carbon nanotube flow sensors, Science

```
{{Citation
| last 1    = Ghosh
| First1    =S
| last 2    = Sood
| First2    =AK
| last 3    = Kumar
| First3    =N
| title     = Carbon nanotube flow sensors
| publisher = Science
| year      = 2003
| id        = {{235-1236}}}}
```

In the next step, we standardized our parsed records into standard and clean string without noise and missing data fields. During the standardization step, each record of our data-set mapped with specific patterns of correct data to make sure all the fields are being defined in the right data format. Figure 2 shows flowchart of preprocessing step in this research.

[7] Patent prosecution is the process of drafting, filing, and negotiating with the US Patent and Trademark Office for the grant of a patent. Patent prosecution includes drafting a patent application and filing the application with the Patent Office.

The preprocessing on NPL data, was one of the main contribution of our research as we were able to extract 27,111 clean standard article NPL records out of set of articles, book items, proceedings, white papers, reports, etc. Extracted NPL articles standardized in separate fields such as title, first author, publisher journal, year, start page, end page, etc. out of NPL raw data. To do this preprocessing step of NPLs we used Java coding in three steps to pars and standardize the data.

Figure 2: Data preprocessing steps

2.2.1.3 Data Preprocessing

Data preprocessing step was one of the most important steps of this research as a step needed to be done before any type of data mining analyses. By preprocessing, we needed to make our data-set clean and free of noise and missing data (Pyle, D., 1999). We needed the pre-processing step to insure that all records are in standard format and we do not have incorrect values in data fields. In other words, in pre-processing step, we standardized out data matrixes to assure the quality of research by using correct standard data.

Data pre-processing includes cleaning, normalization, transformation, feature extraction and selection, etc. (Han, et al., 2011). The main steps of data cleaning and data reduction in our data preprocessing are as follows:

- Ignore the tuple: In our clustering phase, we had to ignore some records as they did not belong to any specific cluster label. We added these records to "other topics" cluster and discard them in our analyses. Records in languages other than English, topics with unreadable characters, etc., were among outliers.
- Correct inconsistent data: Since we were focused on patent and article records, we needed to discard all other types of records to be included in network building phase. We parsed all NPL raw strings to extract article records in well-structured format so that we were able to match them with scientific database in next steps.

- Fill in missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies using multiple databases, data cubes, or files. In this step we double checked all our data records by filtering noisy data such as Null values, duplicates and nonstandard characters using MySQL.

By having preparing the cleaned records of our data-set, the next step of research was to cluster our data records as we will explain in next sub-section.

2.2.2 Data Clustering

In data clustering phase, we applied standard vector space algorithm as the build-in text mining based clustering method using iNSIGHT software tool. This method allowed us to partition articles and patents into subsets called clusters into scientific and technological domains. Text mining performed separately on the article and patent databases in order to distinguish between science domain (article-based) clusters and technology domain (patent-based) clusters, as knowledge diffusion (article network) and technology diffusion (patent network) are quite distinct.

Following the clustering steps, our final step consisted of picking descriptive, human-readable labels for the clusters produced by our document clustering algorithm. It was done by ranking the words appearing in each cluster and finding the most relevant labels best describing each cluster. The two sets of domain clusters (article- and patent-based) were merged separately based on the keyword similarity to see where the science and technology clusters have more similarity and interactions. Figure 3, shows the standard data clustering steps as described.

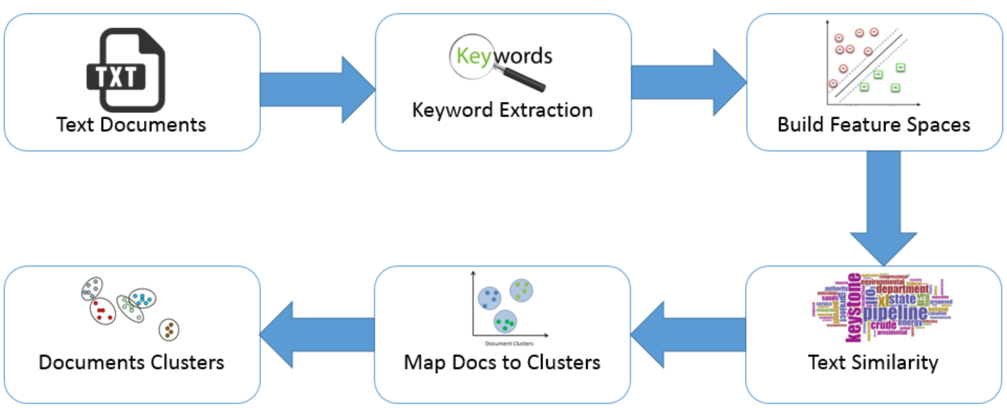


Figure 3: Text Mining Process

In next sub-sections we will discuss some of data clustering methods used by different research papers to cluster and analyses data-sets, while focusing on vector space model as the method we used to cluster our data in this thesis.

2.2.2.1 Cluster Analysis Methods

Data analysis has application in many different computing applications. This technique can be used either in a design phase or as part of the on-line operations. A key element in procedures of data analysis is the grouping, or classification of measurements based on either (i) goodness-of-fit to a postulated model, or (ii) natural groupings (clustering) revealed through analysis. Cluster analysis is the formation of set of patterns into clusters based on similarity. The organization of patterns is usually represented as a vector of measurement or a point in the multidimensional space (Jain, et. al., 1999).

Looking back at the history of patterns from data it goes back to methods of identifying patterns in data which include Bayes' theorem (1700s) and regression analysis (1800s). By the increasing power of computer technology and the massive growth of data size and complexity, we see direct role of computer science in data analysis field, such as neural networks, cluster analysis, genetic algorithms (1950s), decision trees and decision rules (1960s), and support vector machines (1990s). According to Kantardzic (2003), data mining is defined as the process of applying methods to extract hidden data patterns. Figure 4, shows the iterative method in extracting patterns in text mining based clustering.

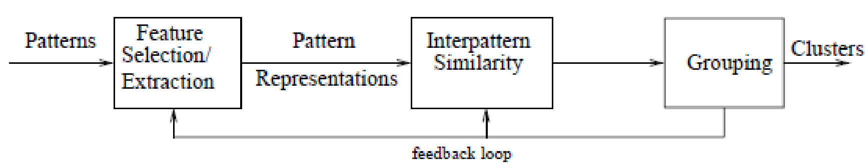


Figure 4: Data Clustering Steps (Jain et al. 1999)

Surveys in data mining and text mining techniques were our main reference to choose the most proper technique in our clustering phase of this thesis. Based on ACM SIGKDD in 2011, data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. Regarding the data mining surveys, “survey of clustering data mining techniques” (Berkhin, 2006) was studied to review various data mining methods which authors used in previous researched in patent and article clustering. In our literature review we focused on methods to be used in article and patent clustering based on keyword detection. Our main survey we used in text mining techniques is a survey by Aggrawal and Zhai (2012). Based on this study, the quality of data mining method any such as classification and clustering is highly dependent on the noisiness of the features that are used for the clustering process.

In choosing the clustering method in this thesis, we used some basic reviews and surveys in general data clustering methods. Surveys of the state of the art in clustering was reported by Dubes and Jain (1980) and Lee (1981), on comparison of various clustering algorithms for constructing the minimal spanning tree and the short spanning path. Cluster analysis was also surveyed by Jain et al in 1986 (Jain et al., 1986), while they reviewed image segmentation by clustering in one more recent research work (Jain and Flynn, 1996). We can also mention comparisons of various combinatorial optimization schemes, based on experiments, which reported Al-Sultan and Khan (1996).

Following up the literature review, we used singular value decomposition to locate terms in each records and to build the terms/phrase document matrix which is conducted to identify the full set of topic names. In the next step, article/patents get assigned to each topic using the standard vector space model while algorithm iterates through the records' set and at each stage various weights / scores are assigned to the topics based on different parameters which we will discuss in the next sections.

As mentioned in research methodology, after data preparation and preprocessing we clustered our data records based on text mining method of vector space model. The main reason for this step was to investigate scientific and technological domains in large citation network which leading citing and cited article/patents belong to. Also we needed to study the evolution of our top scientific clusters over research timeline and see how interaction between different scientific and technological domains takes place. To achieve the mentioned objectives we needed to categorize our records into specific clusters which represent scientific and technological domains.

We used SVD[8] and Standard vector space model as built-in algorithms used in iNSIGHT software tool. The advantages of using the software tool was having the flexibility to tune various aspects

[8] Singular Value Decomposition

of the overall clustering process, as it was possible for us to make the algorithm focus on larger broader topics at one time or finer multi-word topics. Algorithm parameters and their chosen values for this research are discussed in details as we continue on this section.

Discussing the vector space model in more details, according to Dubin (2004), it has advantages such as: being a simple model based on linear algebra, it is based on term weights and also allows ranking documents according to their possible relevance. Following up our literature review on clustering methods, vector space model is a frequently used model in the field of information retrieval, which represents documents as vectors. The entries in the vector correspond to terms in the vocabulary. Binary vectors have a value of 1 if the term is present within a particular document and 0 if it is absent. According to Lee. et.al. (1997), many vectors make use of weights that reflect the importance of a term in a document, and/or the importance of the term in a document collection. For a particular cluster of documents, we can calculate the centroid by finding the arithmetic mean of all the document vectors. If an entry in the centroid vector has a high value, then the corresponding term occurs frequently within the cluster. These terms can be used as a label for the cluster. One downside to using centroid labeling is that it can pick up words like "place" and "word" that have a high frequency in written text, but have little relevance to the contents of the particular cluster. Looking at different text clustering methods, we also had Boolean model as one of other popular methods used in text clustering but comparing these two methods, in Boolean model similarity function is Boolean therefore we can only have exact-match and not partial matches. Also, Retrieved documents in Boolean match are not ranked plus all terms are equally important as Boolean operator usage has much more influence than a critical word.

As we used vector space model as our clustering and cluster labeling algorithm as it examines the contents of the documents per cluster to find labeling that summarize the topic of each cluster and distinguish the clusters from each other. In the next section we will discuss the clustering algorithm in more details and then after we continue by reviewing the steps we took to apply vector space model on data records.

2.2.2.2 Vector Space Model

Based on Raghavan, et al. (1986), the vector space model procedure can be divided in to three stages. First, document indexing where content bearing terms are extracted from the document text. In the second phase, the weighting of the indexed terms is taking place to enhance retrieval of document relevant to the user. The last stage ranks the document with respect to the query according to a similarity measure.

The **tfidf**[9] weight is a weight often used in information retrieval and text mining. This statistical measure is used to evaluate how important a word is to a document in a text. According to Lee, et. al. (1997), the importance of a key word increases proportionally to the number of times it appears in the document but is offset by the frequency of the word in the corpus. Variations of the **tfidf** weighting scheme are often used by search engines to score and rank a document's relevance to a user's query (Figure 5).

The term frequency in the given document is calculated as the number of times a given term appears in that document. [10] To give a measure of the importance of the term ti within the particular document:

Equation 1: Term Frequncy in a document

$$\mathrm{tf} = \frac{n_i}{\sum_k n_k}$$

With ni being the number of occurrences of the considered term, and the denominator is the number of occurrences of all terms (Lee, D., et. al., 1997).

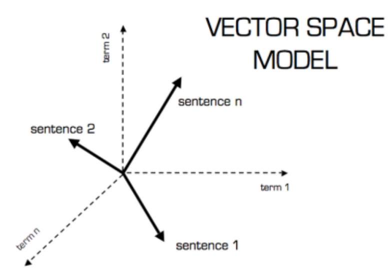


Figure 5: Vector Space Model (Lee, et. al. 1997)

The key steps followed while clustering records are:

1. At the first stage stemming and appropriate stop -words specific to the language of the text is

[9] term frequency inverse document frequency
[10] This count is usually normalized to prevent a bias towards longer documents (which may have a higher term frequency regardless of the actual importance of that term in the document)

Applied to locate terms and phrases. SVD (Singular Value Decomposition) of the terms/phrase document matrix is conducted to identify the full set of topic names.

2. The clustering algorithm then iterates through the document set and at each stage various weights / scores are assigned to the topics based on different parameters (some of these can be set by the user too). Documents then get assigned to each topic using the Standard Vector Space Model. The number of iterations to run and the approximate quantity of topics to be picked up in each iteration, can be influenced via necessary configuration parameters by the user. The filtered list of topics contains those whose weights/score fall above a certain threshold.

3. By hierarchical clustering then after the top level topics are finalized the iterative procedure is repeated for top level topic to identify related sub-topics.

In our research, we calculated each of the parameter values above using iNSIGHT tool to assign each of our records to cluster labels. We needed to set the values of above parameters in a way to have optimum number of cluster labels while considering the cluster size. For example choosing lower merge threshold were ended in bigger clusters and bigger overlap areas while at the opposite point we could have high number of clusters with general labels. We also needed to consider these parameters in a way to have smaller cluster of “other topics”.

We tried to reach the optimum state of considering maximum cluster size, minimum cluster size, and hierarchy depth and merge threshold. As the result we were able to extract 44 scientific and 24 technological clusters with least overlaps. Table 2 in section 3.1 shows the values of these clustering parameters in clustering phase.

Having our data records clustered, the next step to discuss is our article-patent network structure as follows in the next sub-section.

2.2.3 Network Build

In this section we discuss our main citation network structure which will help us to have a better understanding of research objectives and methodology.

Following the research methodology, in network building step, we built a large citation network of nanotechnology articles and patents in which the linkages between the nodes represent citations while the nodes could be either articles, NPL article or patents. Clustering nodes allowed us to cover multiple scientific and technological domains. To identify relationships between clusters while reviewing citation patterns between article and patent nodes, we used this network as an input for the analysis step. The open source software application Gephi was used to build and visualize networks.

Regarding our data-set, we used different citation information to build our main citation network. Our data includes citation information between patents and their NPL articles plus citation information of cited NPL articles. This means focusing on NPL articles as core nodes of our network, considering Figure 6, we have patent nodes as backward citations and article nodes as forward citations of them.

Figure 6: Patent – NPL and NPL – article citation network edge

In NPLs data table, we have a set of patents which are citing a set of articles, this means that they are in patents references, while these NPL articles are citing another set of articles. This means they have other articles in their references which we have their information in our database. In other words, we have ideas in the first level articles (blue nodes of Figure 6) which are cited in the next level articles (green node of Figure 6) and finally they lead to technology domain as they are cited by a patent (orange node of Figure 6). So if we consider -> arrow as **being cited** we will have the pattern of Figure 7 for this network.

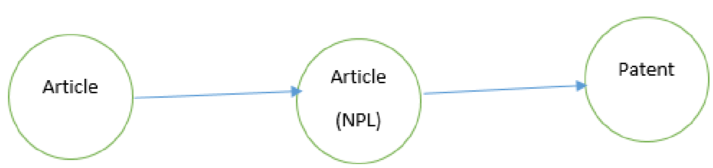


Figure 7: Citation network pattern

Considering this structure, in the next following sections, looking at Figure 8, we can see we have different types of citation links from articles to NPLs and from NPLs to articles they cite. We will discuss these citation relations in next sections.

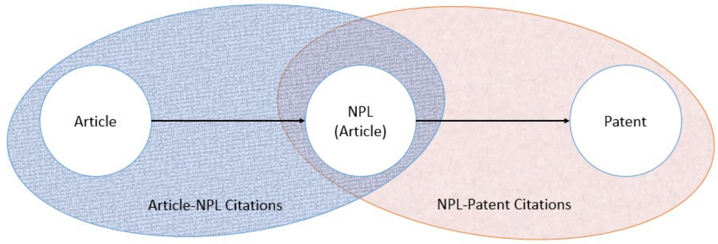


Figure 8: Different citation links between nodes in main citation pattern

We analyzed our patent-article citation network based on various network topological measures which we will review in the next sub-sections.

2.2.4 Networks Topological Measures and Network Analysis

In the next step we did analyses such as cluster analysis, network topological analysis, data trend analysis, and correlation analysis of leading patent/article nodes as well as top main scientific and technological clusters. We also looked into trend and correlation analysis of our main scientific and technological clusters and contribution of top journals and Canadian institutions in development of scientific and technological domains.

As mentioned in section 2.1, some of our main thesis objectives were: to investigate top cited leading articles, find most important (Citing/Cited) NPL, examine most citing patents and their technological domain, study how topological network indicators influence the role of NPL nodes and finally investigate correlation of topological parameters in article-patent citation network. To achieve these objectives, we needed to build the large citation network of articles and patents and then calculate network topological parameters such as degree in/out centrality and betweenness centrality for all nodes belong to largest component of the network. Below we will discuss the relation of each network topological parameter in choosing leading gate-keeper article, NPL article or patent nodes.

As of social network analysis phase of this thesis, here are the main network parameters calculated for network nodes:

1. **Degree Centrality**: Centrality measure shows importance of the network nodes in network information exchange. We can say that a node plays an important role in the knowledge diffusion in the network, when it is widely involved in the communications with other individuals. According to Wasserman and Faust (1994), this kind of involvement is called

the centrality of the vertex. In this thesis the centrality measure that we are more focused on is degree centrality which measures the number of nodes that are directly connected to this node. Clearly, the more a network is connected to other nodes, the more active it will be in the sense of information transfer and consequently, in a way that it will be more central. Generally degree of centrality is defined based on equation 2 as below:

Equation 2: Degree of Centrality

$$DC = \frac{Variation\ in\ Degree\ Centrality\ of\ Vertices}{Maximum\ Variation\ in\ Degree\ Centrality\ of\ Network\ of\ the\ same\ Size}$$

Degree of centrality can be discussed as in-degree or out-degree depending on number of directed incoming or outgoing edges which are connected to a node.

1.1 **Degree of Centrality (in-degree)**: Measures the total number of incoming edges connected to a node in a directed graph. In this research, the higher this value is, the more node cites other nodes (It is using more references). Itis has a wider scope and has a more inter disciplinary role in research.

1.2 **Degree of Centrality (out-degree)**: Measures the total number of outgoing edges connected to a node in a directed graph. In this research, the higher it is, node is being cited more. It means this node is more used as reference by other article/patents.

1.3 **Betweenness Centrality:** This measure evaluates the significance of a node as a connector between two other nodes that can enhance the knowledge exchange between them.[1]Betweenness centrality takes into consideration the role of intermediary articles or patents, i.e. the articles that lie on the paths connecting two nodes (Wasserman and Faust, 1994). So we can say, the betweenness centrality of a node is defined as the proportion of all shortest paths between pairs of other nodes that contain this node (De Nooy et al. 2005).

1.4 Gate-keeper node: We say that a node k is a gatekeeper if, for some other distinct nodes i and j, k lies on every path between i and j (Easley and Kleinberg, 2010). In other words, A is a gatekeeper because it lies on every path between two subnets.

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

Equation 3: Betweenness Centrality

$$BC = \frac{Variation\ in\ Degree\ Centrality\ of\ nodes}{Maximum\ variation\ in\ Betweenness\ Centrality\ of\ nodes\ in\ a\ network\ of\ the\ same\ Size}$$

In this thesis we calculated betweenness centrality measure of nodes in giant component of our networks using Gephi tool. Giant component is the largest connected component of the graph which contains a fraction of all network nodes and edges. In this case we calculated BC over giant component as we needed to look at connected citation paths from an article which ends to a patent. Considering our data set structure and steps we took to build the network, giant component of our network included 98% percent of all nodes, and the 2% outlier could be discarded. The higher is BC value for an NPL, the higher influence it has on the transfer of items through the network from knowledge to technology domain.

2. Average of citations per article: Citations per paper (sometimes called "impact") is computed by dividing the sum of citations to some set of papers for a defined time period by the number of papers (paper count). The citations per paper score is an attempt to weight impact in respect to output, since a greater number of publications tends to produce a greater number of citations.

2.2.5 Results Interpretation

As the final step, analysis results were interpreted and discussed on various plots and tables. Figure 9 shows a big picture of whole research process flowchart.

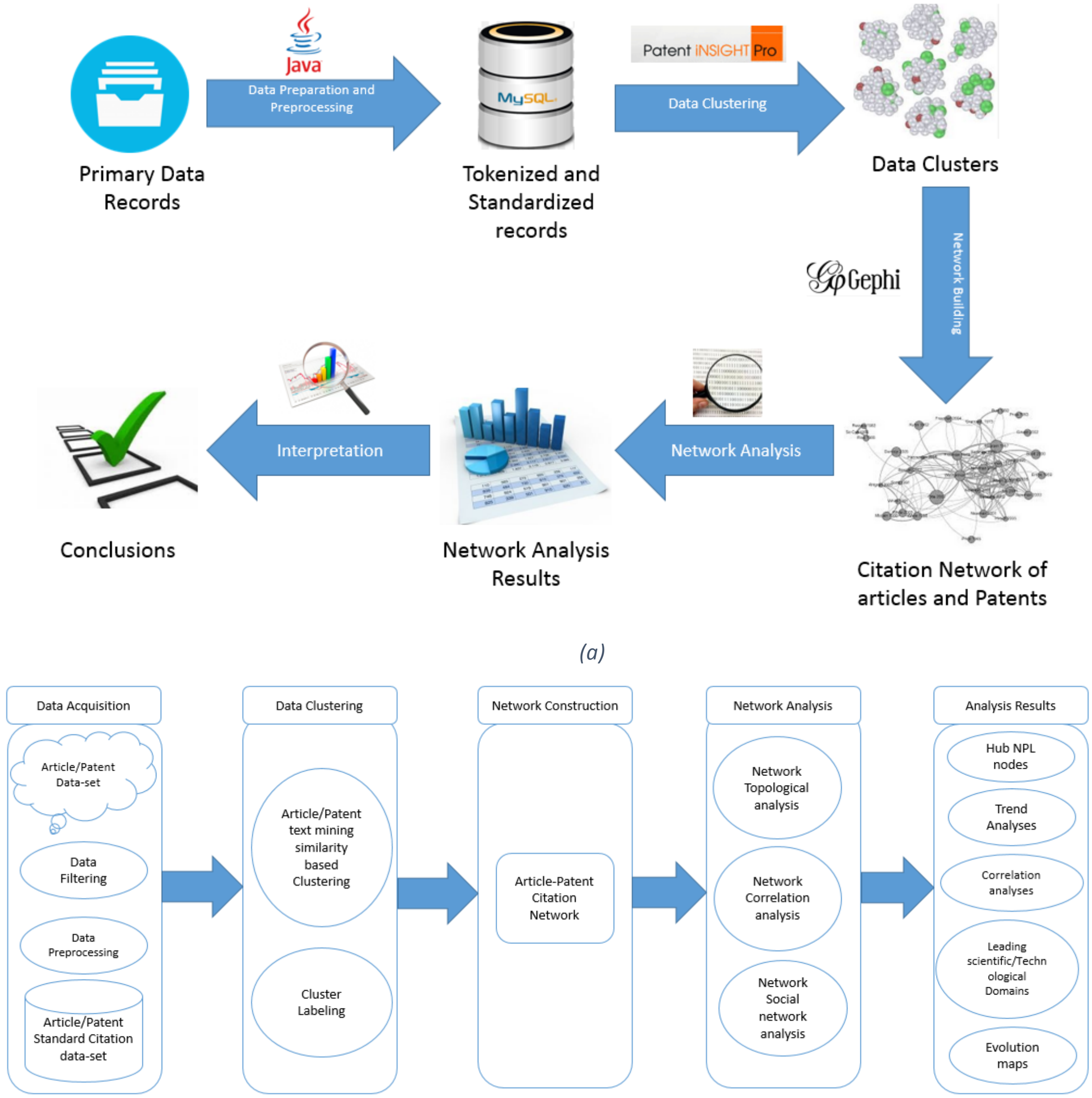


Figure 9: (a) Research methodology flow map; (b) Research Methodology flowcharts

2.3 Chapter Summary

In this chapter we discussed our research objectives, and then research methodology as well as our research flowchart. We also explained our patent-article citation network structure, cluster analysis methods and finally network topological measures which we will use in the next chapter in data analysis phase.

3. Analysis of Data

Introduction

In this chapter we will discuss analysis steps such as cluster analysis, network topological analysis, data trend analysis, and correlation analysis of leading patent/article nodes as well as top main scientific and technological clusters. First of all, we will discuss the big picture of all patent-article citation network and we will continue the discussion by data clustering method and key parameter values we used in clustering process. In the next sections we will look at trend and correlation analysis of our main scientific and technological clusters and contribution of top journals and Canadian institutions in development of scientific and technological domains.

To have a clear view of this chapter, first of all we are going to discuss the big picture of all patent-article citation network while considering the flow of citations and ideas in network. Looking at Figure 10, we visualized the schema of patent-article citation network while categorizing each set of nodes in a separate layer. Since we ran clustering algorithm on different sets of data records, we can see a various scientific/technological clusters and different tiers of nodes in citation network.

NPL nodes are more of our interest in this research as they are playing gate keeping role in transfer of ideas from scientific domains to technological fields. It means that an idea flows from article scientific domains to NPL scientific domains while an article is being cited by another article. This idea is flowing to technology domains when it is being used in a patent. Our evidence for such a citation relation is when a patent cites an article in its references. This is the main idea visualized in Figure 10 which shows how scientific and technological domains interact via flows of citations and ideas and how top technological scientific and technological clusters are contributing in this type of interaction. As we also explained our detailed citation network in section 2.4, and as we can see in Figure 10 the citation flow in this network starts with a set of nanotechnology patents which cited some scientific work that they used in their invention process. This is how we are going to follow the flow from patents to NPLs cited by these patents and then afterwards we continue this flow by identifying some further prior scientific knowledge in articles which NPLs cited. Looking at Figure 10, we can see top clusters of each node set such as patents, NPLs and articles. These clusters are extracted based on the clustering step of methodology and compared with other clusters in terms of number of nodes, number of total citations and citations per articles/patents. The clusters which are mentioned in Figure 10 are based on the results of analyses in this Chapter which we will discuss in the next sections. As mentioned before, the main purpose of pointing to these big picture schema in the beginning of Chapter 3, is to have a better idea of the analysis objectives in this chapter.

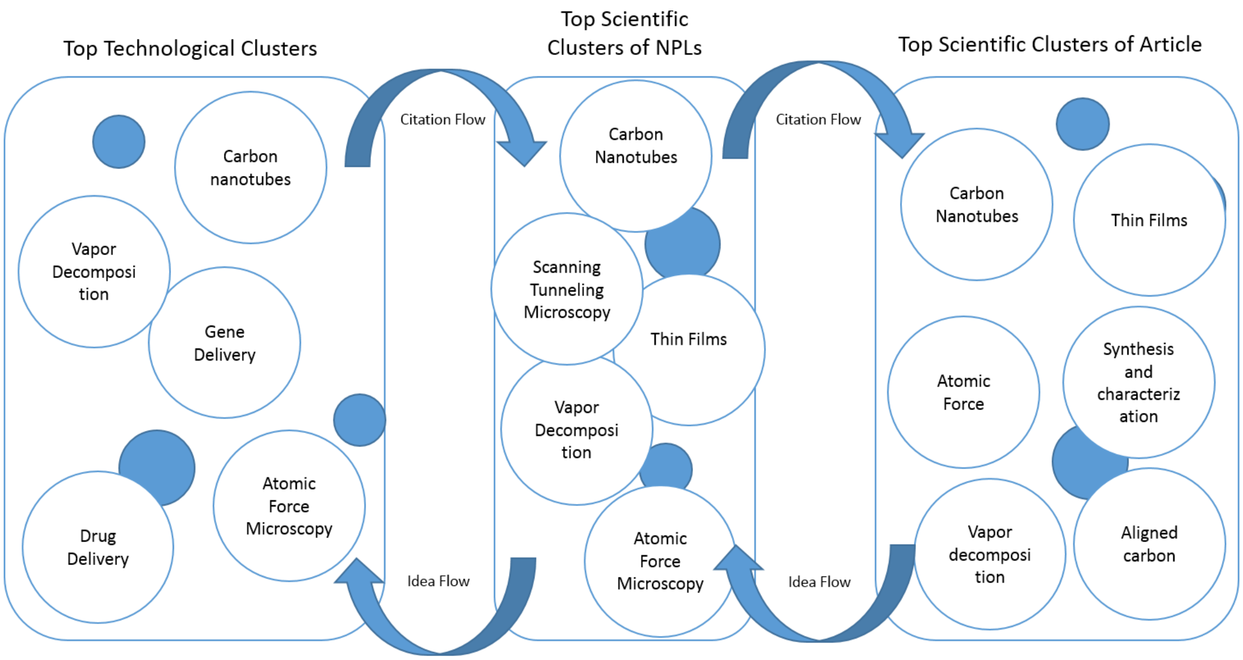


Figure 10: Citation network idea/citation flow schema

3.1 Text Mining Based Data Clustering

One of our research objectives was to investigate top scientific and technological clusters in different network maps to have a better view of clusters overlaps and sizes. Using space vector model which discussed in Chapter 2, we were able to extract 44 different clusters which we will discuss in more details in the next sections. Complete list of clusters is provided in Appendix-A.

As was discussed in research methodology, cluster labeling algorithm was applied iteratively on data in order to extract the clusters' boundaries in the finest state with the least overlap between technological domains.

The optimum state of data clustering is shown in Table 2 according to algorithm parameters. As mentioned each parameter value is chosen based on several algorithm runs so that we were able to see clusters' borders clearly defined with least number of overlaps between them. Also, we needed to set parameter values in a way to have less number of records in “other topics” cluster which needed to get discarded.

Table 2: Clustering parameters values in clustering

Clustering Parameter	Value [Range]
Hierarchy depth	2 [1-5]
Cluster count	50 [1-50]
Maximum top level clusters	10 [1-10]
Maximum sub level clusters	7 [1-10]
Merge Threshhold	0.7 [0-1]
Minimum cluster size	0 [0-1]
Maximum cluster size	0.05 [0-1]
Minimum label lengh	2 [0-8]
TF/DF ration label score weight	0.6 [0-1]
TF label score	0.2 [0-1]
Work Count Score Wight	1 [0-1]

According to Gridlogics Technologies (2014), in applying clustering algorithm, there are some specific parameters which help us to extract clusters. We will continue this section by discussing the parameters in Table 2 and the values chosen for each of them to extract data clusters:

- **Hierarchy depth:** The maximum number of cluster levels to create. Here we can see “hierarchy depth” value chosen as 2 so that we could extract clusters in 2 levels. In this case, for example for cluster “carbon nanotubes” in level 1, we had sub clusters such as “nanotube films”, etc. This two level clustering help us to have a closer look at sub-domains of each scientific and technological domain and their interactions in future studies.
- **Cluster count:** The number of clusters discovered in each clustering pass. The higher the value of this parameter, the larger the total number of clusters. Cluster count value is set based on the number of records we needed to cluster and the size of clusters which we were

planning to have. On this case we chose 50 for cluster count and discarded clusters belonging to “other topics” and other unrelated topics.

- **Maximum top level clusters:** Maximum number of clustering passes to perform on top hierarchy level. With the lowest value of this parameter, the clustering engine will discover only the largest clusters, while with higher values, smaller and more specific clusters will also be created. This was why we set this value as the maximum value of 10 to find more specific clusters.
- **Maximum sub level clusters:** Maximum number of clustering passes to perform on sub-clusters. With the lowest value of this parameter, the clustering engine will discover only the largest clusters, while with higher values, smaller and more specific clusters will also be created. For this value, we chose the parameter as 7 instead of maximum value. The reason was regarding the number of sub clusters which were decreasing a lot by setting the value as maximum. We needed to have a medium level of sub clusters so that the land scape graph of all scientific clusters was not getting unclear.
- **Merge Threshhold:** Low values of this parameter will cause the clustering engine to eagerly merge clusters, which will create larger clusters in which some documents may be irrelevant. High values of this parameter will cause it to merge clusters rarely, which will result in large numbers of small clusters with more relevant documents. We set merge threshold as 0.7 as we needed to have larger number of smaller clusters with more relevant documents.
- **Minimum cluster size:** Determines the minimum allowed size of a cluster in relation to the parent cluster size. E.g. a value of 0.4 means that clusters must not contain less than 40% of the parent cluster's documents (of all documents in case of top-level clusters). We set this value as 0 as we didn’t want to put any constraints on the number of records in level 2 clusters in regards to higher level clusters.
- **Maximum cluster size:** Determines the maximum allowed size of a cluster in relation to the parent cluster size. E.g. a value of 0.4 means that clusters must not contain more than 40% of the parent cluster's documents (of all documents in case of top-level clusters). According to the explanation of previous parameter, we set this value to 0.05 so that we could have higher number of specific clusters.

- **Minimum cluster label:** Determines the minimum label length in words. Labels consisting of fewer words will not be generated. Setting the minimum label length to some higher value (e.g. 4 or 5) may create more specific clusters. In our case we set the value as “2” so that we created maximum 2 word cluster labels.
- **Title word label score:** Assigns higher scores to labels that contain word that appeared in input documents' titles. We set this value as maximum of 1 as we needed to assign higher scores to related records.
- **TF/DF ration label score weight**: Assigns higher score to more general/shorter labels. This value assigned to 0.2 as did not want to have more general short cluster labels. Our target was to follow more specific and exact cluster labels.
- **TF label score:** Assigns higher scores to labels with higher Term Frequency (TF). According to the same explanation for TF/DF ratio value, this value also set as Maximum to give more scores to topics with higher term frequency.

Continuing the discussion about the scientific clusters, here in Figure 11, we are going to discuss first top 5 ranked scientific clusters in number of article records, and we will continue by looking at their correlation and evolution over research timeline.

First top 5 ranked scientific clusters in number of article records are as below:

- Carbon nanotube
- Thin films
- Vapor decomposition
- Atomic force microscopy
- Scanning tunneling microscopy

A cluster map contains visualized information of sets of categorized objects. Its main purpose is to show if and how these sets overlapped. Here as we can see in Figure 11, clustering map shows that the greatest overlap exists between “atomic force microscopy” and “scanning tunneling microscopy” clusters. The smaller overlap we have between two clusters the lower number of common records between them. It means clusters with larger overlap areas are more related to each other and can have more interactions in terms of common articles or patents. These scientific clusters are extracted from giant component of Canadian nanotechnology citation network of articles[11]. Looking at the Figure 11 it is clear that both “carbon nanotubes” is the biggest scientific cluster with the highest number of nodes while it has overlaps with other top clusters which keeps a considerable number of nodes interconnected in the network. The higher overlap between “atomic force microscopy” and “scanning tunneling microscopy” clusters is expected as both domains are related to microscopy. Both these techniques are used to form images of surfaces using a physical probe that has the ability of scanning nano-particles.

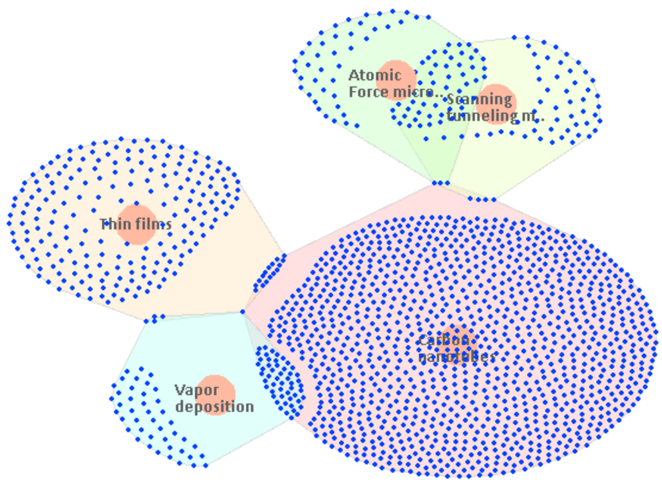
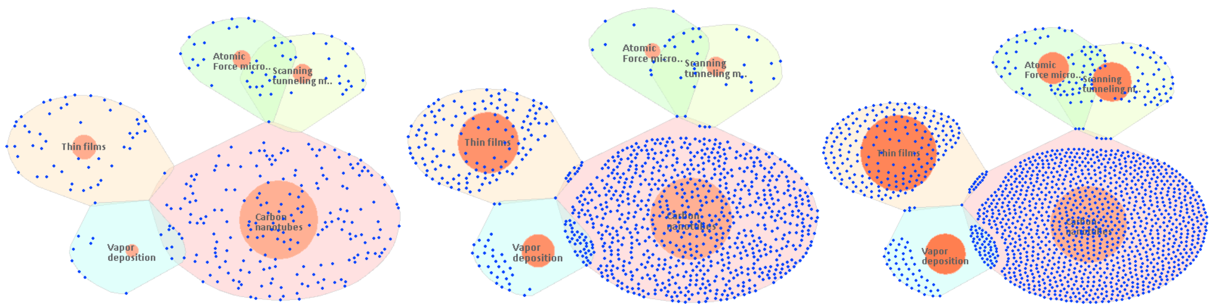


Figure 11: Scientific NPL articles' top 5 clusters maps

Figure 12, shows the growth in number of records for each of our top scientific clusters in NPLs. In the first period from 1995 to 1999, we can see “carbon nanotubes” as the cluster with higher number of articles, while other clusters such as “vapor decompositions” and “thin films” did not grew much. This is justified according to nanotechnology timeline literature since “carbon nanotubes” have been interesting to scientists since 1951, while some other fields such as “thin films” were more in focus of research works as of 1991. Looking at second period from 2000 to 2004, we can see “carbon nanotubes” and “thin films” have faster growth but still clusters related to microscopy fields are growing in a lower rate. This can be related to the fact that, these scientific fields founded by the invention of scanning tunneling microscope in 1981, and they started to grow more since gradually scanning tunneling microscopes got more advanced as the main tool in these two areas. In period from 2005 to 2008, we see growth in all five top clusters which shows higher

[11] As mentioned before, giant component is the largest connected component of the graph which contains a fraction of all network nodes and edges. We extracted these clusters over giant component as we needed to look at connected citation paths from an article to a patent.

rate of scientists' interest in these nanotechnology related domains in recent years and various devices are developed over using nanotechnology knowledge.



(a) 1995-1999 (b) 2000-2004 (c)2005-2008

(b) Figure 12: Clusters growth map

In our correlation map of Figure 13, nodes are connected based on common articles which are shared between the nodes. Two nodes with a high number of common nodes will have a thicker, darker line connecting them. In Figure 13 we observe stronger correlation between two sets of clusters which are "atomic force microscopy" and "scanning tunneling microscopy" and also "carbon nanotubes" and "vapor decomposition". In 2008, a similar kind of analysis was done as content map analysis on nanotechnology USPTO patents of years 2002 to 2004 (Chen, 2008). Looking at Chen's findings, we can see "carbon nanotubes", "thin films", "atomic force microscopy", "force composition" among the top main content map clusters. But here in our research, we tried to map the correlation between different clusters besides showing the clusters mappings. This can help us understand the interactions between main scientific clusters and changes in level of interactions between nanotechnology science topic areas, following up our objective on investigation of main scientific domains. Comparing our results with Chen's (2008) we did data clustering over a wider timeline using more data records. Moreover, according to his results we can see his strong correlation between "carbon nanotubes", "thin films" and "atomic force microscopy", which is also confirmed by our results.

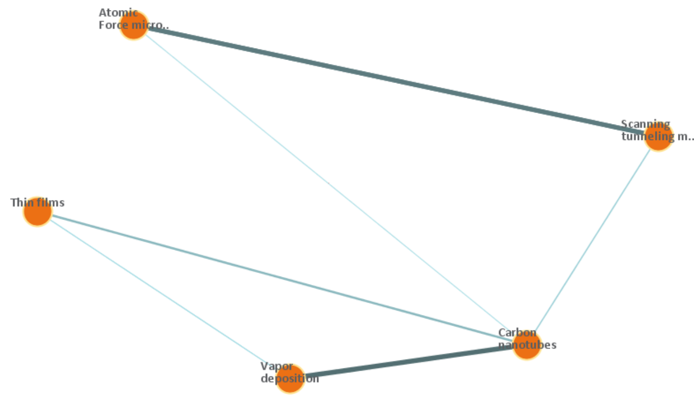


Figure 13: Top scientific articles' clusters correlation diagram

Following up the discussion on development of scientific clusters, we extracted correlation map for all our NPL scientific clusters in regards to their publication year. As Figure 14 shows Canadian NPLs' scientific clusters correlation map for years 1995 to 2008, it also confirms "carbon nanotube" cluster with highest number of records in common with other clusters and also most correlated node comparing to other scientific clusters in our network. Looking at year nodes around the network we observe that we have the highest number of publications between 2002 till 2008 comparing to number of publications in research timeline before 2004. This was expected as we got higher number of publications in the recent 4 years of our timeline comparing to earlier before 2004. According to our literature, other research studies on nanotechnology data were mostly focused on shorter time lines similar to research by Li and his colleagues (2008) on patents. Another interesting finding from Figure 14, is that we can see more connections from year nodes to different scientific cluster nodes in recent years. This shows foundation of more fields in nanotechnology as we can see the number of connected domains to year nodes after 2002 is increasing. It also can be explained by the fact that number of our NPL records in our data-set started increasing significantly after year 2002. This is also confirmed by literature, while DOE National Laboratories (2007), by the mid-2000s new scientific attention began to flourish in nanotechnology. According to productive nano-systems technology roadmap report (2007), projects emerged to produce nanotechnology roadmaps which center on atomically precise manipulation of matter and discuss existing and projected capabilities, goals, and applications.

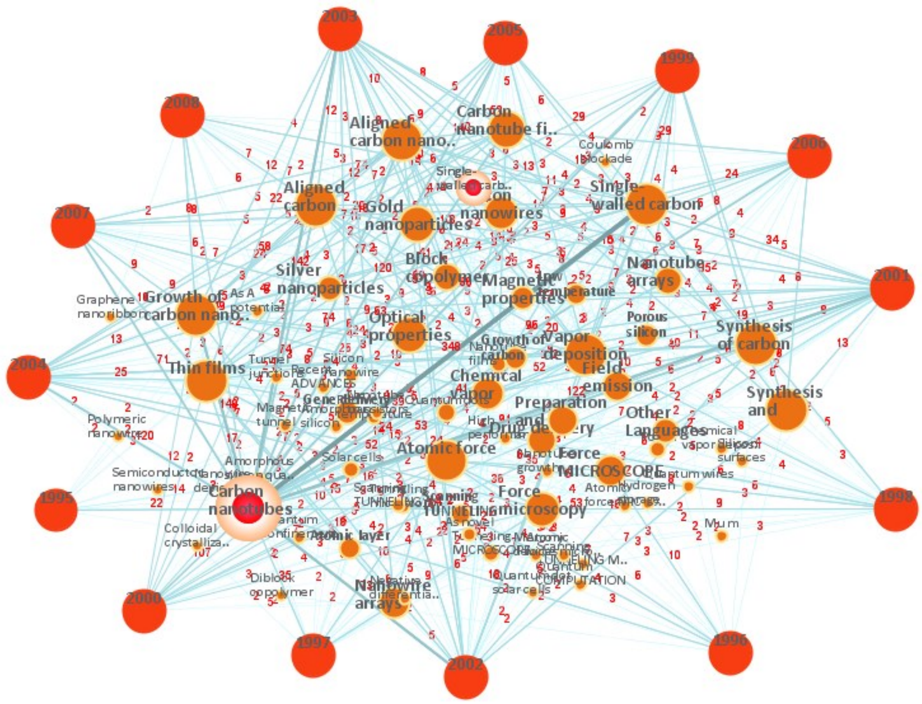


Figure 14: Canadian NPLs' scientific clusters correlation map 1995 - 2008

3.2 Patent – NPL-Article Citations

In the following sub-sections, we are going to discuss topological analysis of our patent-article network regarding different types of citation links between nodes, such as: NPL-article citations, patent-NPL citations and so on.

3.2.1 Network Topological Analysis of NPL-Article Citations

Having a closer look at NPL articles data, we can see distribution of all records in plot of Figure 15. Regarding our research timeline, we chose our patents between years 1995 till 2008, and accordingly the NPL articles which were being cited by these patents were chosen based on the same timeline. It shows number of NPLs published in each year in an accumulative trend plot. As we see, number of articles published in nanotechnology fields has an increasing trend.

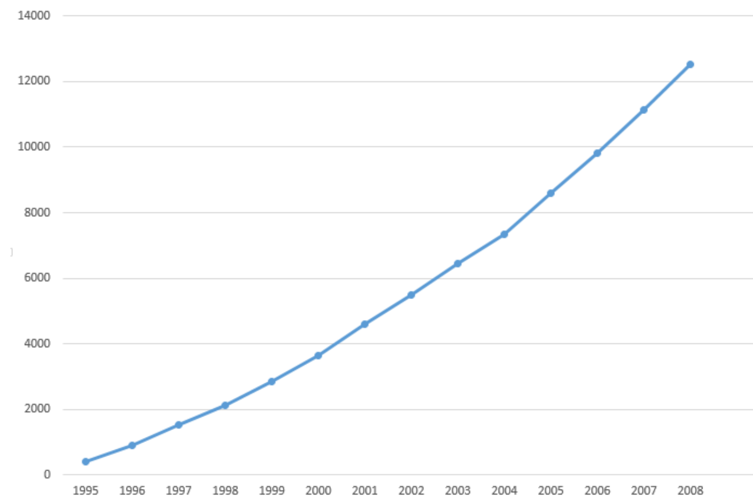


Figure 15: Overall NPLs Publication Data Trend (Cumulative)

As we discussed the patent-article citation network structure, NPLs are considered as middle tier bridge nodes between our first tier which are patents and third tier as articles. We tried to show the gate-keeping role of NPLs in Figure 16, as it shows a big top view of our large citation network with NPL nodes in red color while patent nodes are colored in green and all other cited articles in blue. We visualized nodes in different colors and size of them base on in-degree value to highlight the role NPL nodes with higher value of in-degree, as the bigger a node size is the higher in-degree value it has. According to our discussion before, in-degree value of an NPL article is important to us as it shows the number of nodes which are citing that specific article. It means that bigger sized nodes in Figure 16, are connected to more patents from technological cluster of our network and they are connecting more nodes from technological clusters to scientific clusters. Having a closer look at network of Figure 16, we can see most of nodes with high in-degree value are placed near left bottom of the network which is related to our most cited domain as “carbon nanotubes”. This was expected since we have “carbon nanotubes” as the cluster with higher number of articles and stronger connections to other clusters. In the next sections of this chapter we will also see each scientific cluster mapping on the same network. We will continue this discussion by looking at the most citing and cited NPL articles in our patent-article network following up one of our main objectives of this thesis on investigation of leading NPL gate-keeper nodes in connecting technological and scientific domains of our citation network.

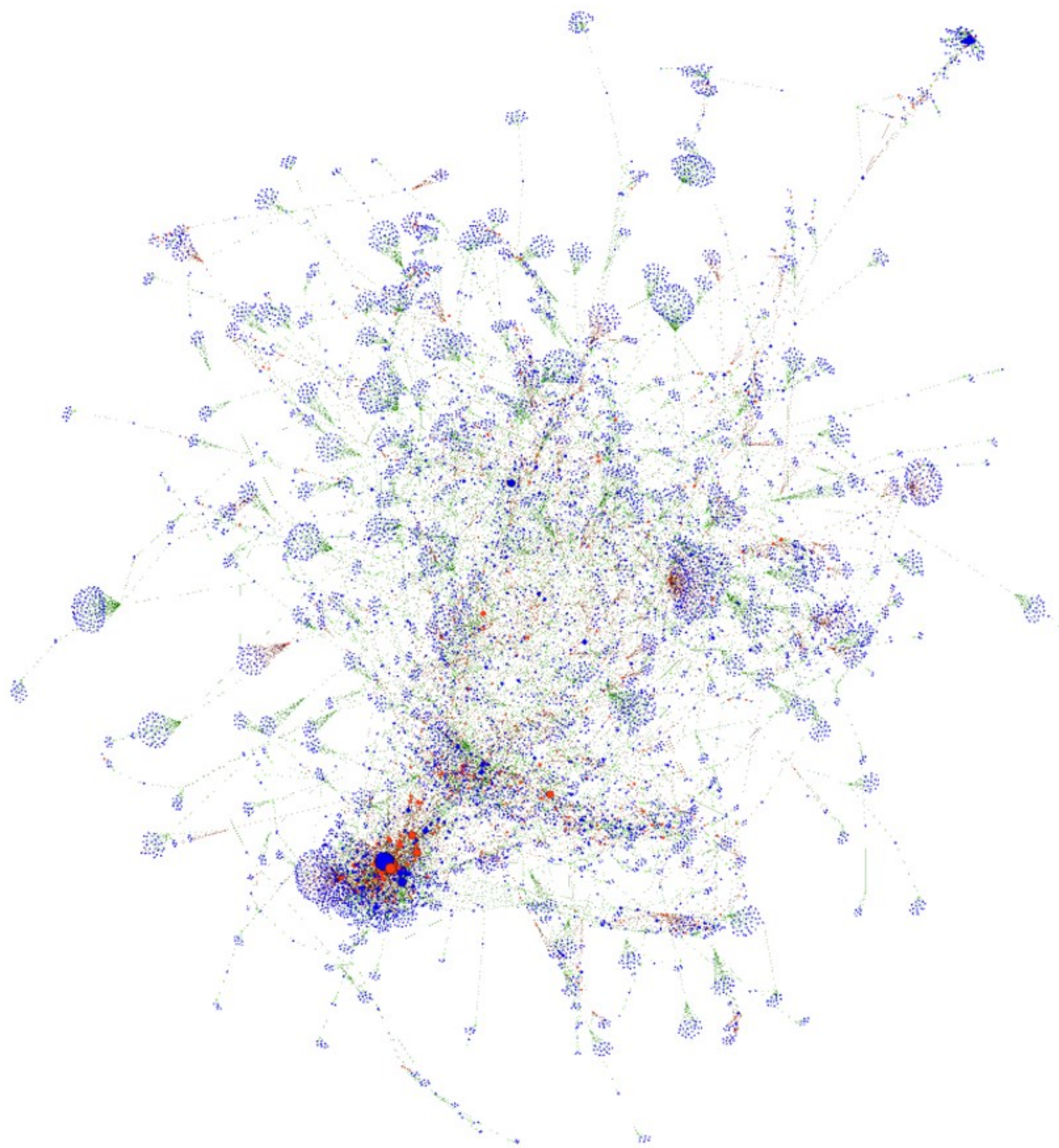


Figure 16: NPL nodes with higher in-degree value in red in comparison with non-NPL blue nodes in networks' giant component

Following up the discussion about most important NPLs, most citing NPL articles are shown in Table 3. These NPLs are sorted based on their value of out-degree parameter which shows the number of other articles they cited. For each of these NPL articles we have related year, scientific domain (based on scientific data clusters as of section 3.1) and their publishing journal. Table 3 helps us to have a top view of top main citing NPL articles which keep technological and scientific clusters connected by citing other articles. These top citing NPLs are important to us as they keep more articles from scientific clusters related to technological domains through NPL gate-keeper nodes. Moreover, the higher out-degree of an NPL node is, the more it cites various articles, which shows the bigger scope which these articles cover in nanotechnology. It means NPLs of table 3 have the highest number of connections to scientific domains.

Table 3: Top citing NPLs (Canadian articles)

Rank	Out-degree	Year	Scientific domain(s)	Journal
1	349	2000	Nanowire arrays	CHEMPHYSCHEM
2	144	2005	Gold Nanoparticles	ENVIRONMENTAL HEALTH PERSPECTIVES
3	96	2007	Magnetic Fields/Thin Films	JOURNAL OF DISPLAY TECHNOLOGY
4	96	2003	Organic structures	TISSUE ANTIGENS
5	94	2000	Atomic Force Microscopy/Carbon Nanotubes	SCIENCE
6	62	2004	Carbon Nanotubes	MRS BULLETIN
7	60	2002	Drug Delivery	ATHEROSCLEROSIS
8	56	2007	Optical Properties/ Gold Nanoparticles	ATHEROSCLEROSIS
9	52	2000	Scanning Tunneling Microscopy	EMBO JOURNAL
10	50	2006	Nanoscale/Composite Material	CURRENT ANALYTICAL CHEMISTRY

In Table 4, top cited NPLs are sorted based on their value of in-degree parameter. Similarly as in Table 3, for each of these NPL articles we have related year, scientific domain and their publishing journal. Table 3 shows top main NPL articles which were cited the most by other nodes in patent-article citation network and established a strong connection between technological and scientific clusters. Results of Table 4 helps us in investigation of leading cited NPLs in our article-patent network. These cited articles are important to us as they include some prior knowledge which inventors used in their inventions.

It other words, NPL articles of Table 4, had the greatest influence on technological and scientific domains because they were cited by higher number of patents and articles. We also mentioned the related journal of each leading cited NPL node to have more information of their publishers.

Table 4: Top cited NPLs (Canadian articles)

Rank	In-Degree	Year	Scientific Domain	Journal
1	113	2006	Carbon Nanotubes /Nano Structures	Nature
2	113	2005	Vapor Decomposition/ Scanning Probe Microscopy / Carbon Nanotubes	Nature
3	112	2005	Composite Materials	JOURNAL OF APPLIED PHYSICS
4	109	2005	Optical Properties	LANGMUIR
5	101	1996	Carbon Nanotubes	PHYSICA C
6	98	2003	Composite Materials/Thin Films	CHINESE PHYSICS
7	79	2001	Scan Tunneling	IEEE TRANSACTIONS ON APPLIED SUPERCONDUCTIVITY
8	55	2002	Aligned Carbon	JOURNAL OF APPLIED PHYSICS
9	54	2005	Carbon Nanotubes	APPLIED PHYSICS LETTERS
10	43	2005	Silver Nano Particles	LANGMUIR

Following up the discussion on most cited NPL articles in patent-article citation network, we calculated “average number of citations per article” for each of our scientific domains in NPLs set. Average number of citations per article is also used as one of citation metrics in other research papers such as study by Viera and Gomes (2009). They used this value as of citation metrics in analysis of articles impact. According to their results, they found a linear behavior between the citation per article and impact factor and for Mathematics and Physics results showed near to the linear behavior. In our case, for each of scientific domains we calculated summation of in-degree value of all article nodes and then divided it by the number of articles as a metric to see how different scientific domains influenced technological domains. Figure 17 shows “scanning probe microscopy” as the scientific domain with highest value of citations per article average. This also shows high rate of this cluster's impact on technological domains as well as the significance of highly cited articles in this field. Considering the nature of "scanning probe microscopy" field, we can realize significance of articles in this field to development of the inventions which end up patents as it gives inventors the ability to measure small local differences in object height while it does not require a partial vacuum but can be observed in air at standard temperature and pressure or while submerged in a liquid reaction vessel (Fritz, et. at, 1994). The higher rate of citations of "scanning probe microscopy" is also justifiable regarding the nature of this field as we have high number of articles in other clusters such as "atomic force microscopy" and "scanning tunneling microscopy" related to this area.

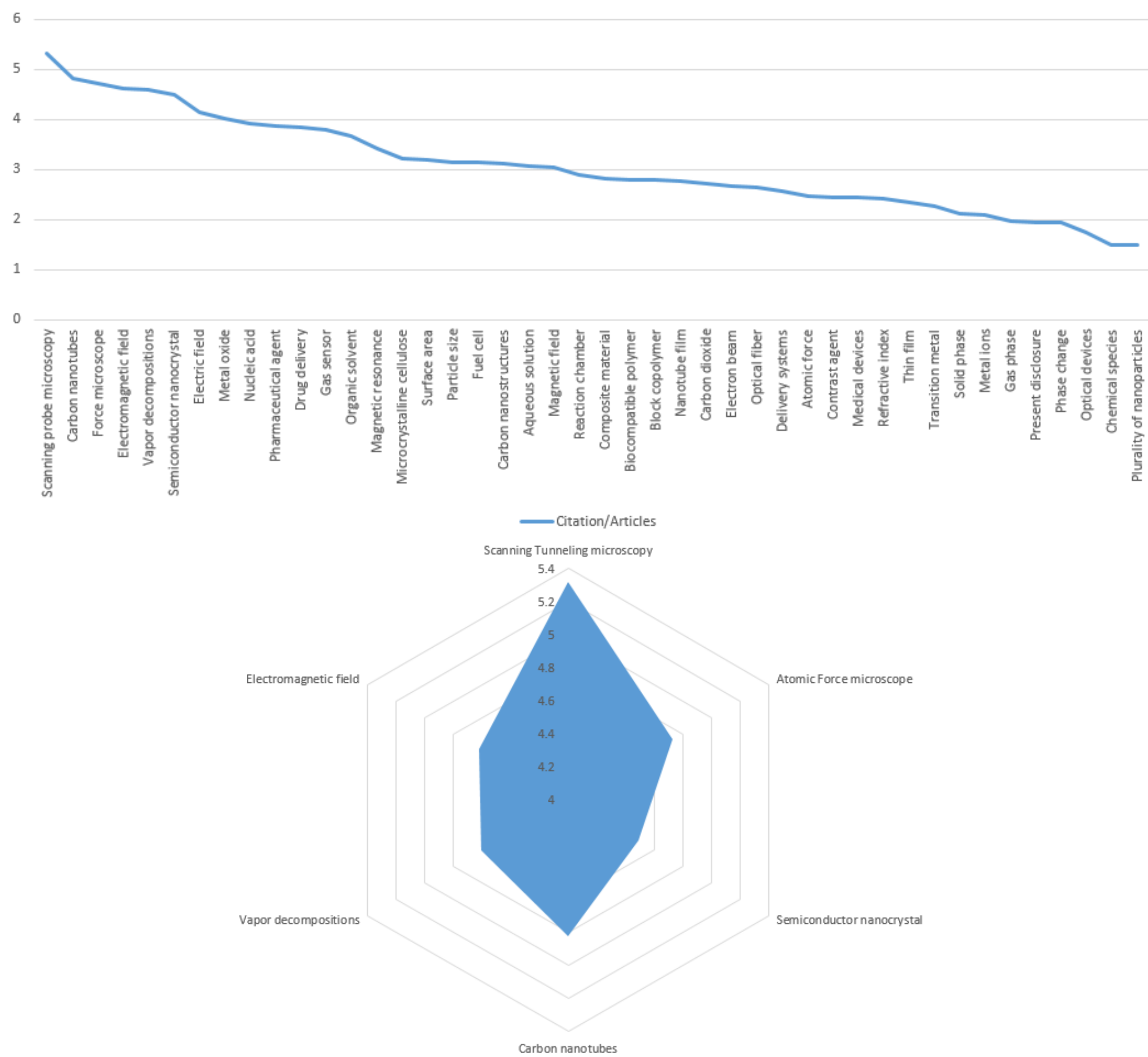


Figure 17: Citations per article in scientific clusters

In Table 5, top citing NPLs are sorted based on their value of betweenness centrality. Table 5 helps us to have a big picture of top main NPL articles which play gate-keeper role in connecting more patent article nodes. Looking at Figure 18, direct correlation between degree centrality and betweenness centrality is investigated as the rising trend of betweenness centrality while in-degree value grows.

Table 5: Top NPLs in betweenness centrality

Rank	Betweenness Centrality	Year	Domain	Journal
1	2.78E-05	2000	Carbon Nanotubes	PHYSICAL REVIEW B
2	8.50E-06	2000	Nanoparticle Arrays	CHEMPHYSCHEM
3	6.18E-06	2002	Nanostructures/Magnetic Fields	PHYSICA C
4	3.64E-06	2005	Composite Materials	JOURNAL OF APPLIED PHYSICS
5	3.35E-06	2005	Carbon Nanotubes	ENVIRONMENTAL HEALTH PERSPECTIVES
6	3.09E-06	2005	Silver Nano Particles	LANGMUIR
7	2.88E-06	2005	Optical Properties	LANGMUIR
8	2.75E-06	2002	Semi-Conductors/Thin Films	PHYSICA C
9	2.51E-06	2005	Vapor Decomposition/ Scanning probe microscopy	Nature
10	2.50E-06	2003	Composite Materials/Thin Films	CHINESE PHYSICS

Following up discussion on relation between in-degree centrality and betweenness centrality values in our patent-article citation network, in Table 5, we created the co-relation diagram of betweenness centrality and in-degree centrality for top 200 article nodes of our citation network with highest in-degree as of Figure 18, to investigate if there is such a relation between these two values. Results of this chart show a slightly increasing trend for degree centrality for 200 nodes sorted decreasingly for in-degree value. In other words, while we sorted a set of nodes based on one of their in-degree value centrality of betweenness' value also followed an increasing trend. This result will confirm what Yan and Ding (2010) stated regarding the correlation between different centrality parameters. They used co-authorship data from 18 journals in the field of library and information science (LIS) with a time span of 20 years (1988–2007). As Yan and Ding state in their research article, there is a significant correlation between these two parameters as citation is a metric of article impact, and centrality is a metric of author impact, so it is not surprising to find that they are correlated but also differ in their representation (Yan and Ding, 2010).

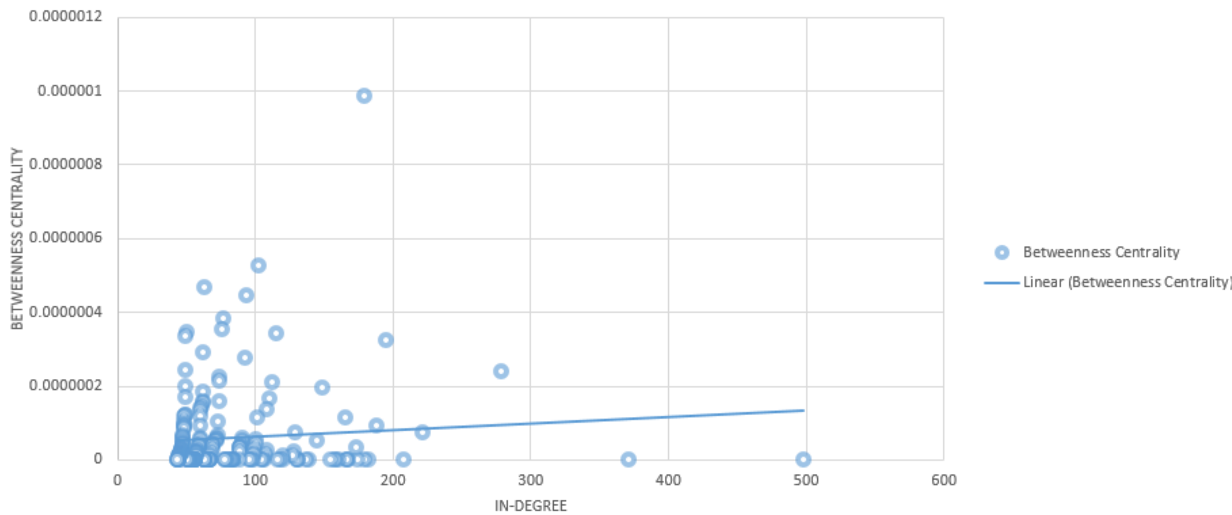


Figure 18: Correlation between citation in-degree count and betweenness centrality

Comparing our results to Yan and Ding's research, as mentioned before they used a co-authorship data of international journals in the field of library and information science while we used citation data from citation network of nanotechnology articles and patents. Our results are interesting in a way that they show the positive correlation between betweenness centrality and in-degree value of nodes in a citation network while the same type of positive correlation was discussed as results of Yan and Ding on co-authorship networks. Moreover, according to Yan and Ding centrality measures can be useful indicators for impact analysis, and regarding the similarity between our research and theirs, we would be able to use betweenness centrality as an indicator of journals' impact in our network in future studies.

3.2.2 Network Topological Analysis of Patent-NPL Citations

In this section, considering the citation pattern in Figure 19, here we are going to discuss patent-NPLs article citations. To do so, we are going to take a closer look at top citing patents in Canadian nanotechnology citation network as these patents play a key role in keeping technological and scientific clusters of the citation network connected.

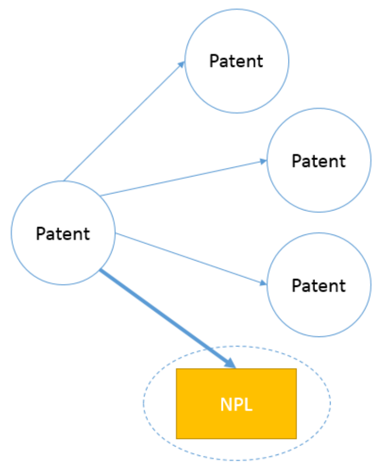


Figure 19: Patent-NPL citations

Table 6: Top citing patents

Number	**Out-degree**	**Year**	**Technological Domain(s)**	**Assignee**
1	36	2007	Carbon Nanotubes	Massachusetts Institute Of Technology, The Brigham And Women's Hospital, Inc.
2	31	2009	Carbon Nanotubes/Force Microscopy	Institute National De La Recherché Scientifique
3	28	2003	Optical Properties	D-Wave Systems, Inc.
4	27	2001	Quantum Computing	D-Wave Systems, Inc.
5	27	2003	Electric field/ Quantum Computing	D-Wave Systems, Inc.
6	24	2008	Magnetic properties/ Electric Field	D-Wave Systems, Inc.
7	24	2002	Quantum Computing	D-Wave Systems, Inc.
8	22	2006	Electric field/ Quantum Computing	D-Wave Systems, Inc.
9	21	2006	Gold Nanoparticles	Raimar Loebenberg, Finlay Warren H, Roa Wilson H, Leticia Ely
10	21	2002	Atomic Force Microscopy/Scanning Probe Microscopy	D-Wave Systems, Inc.

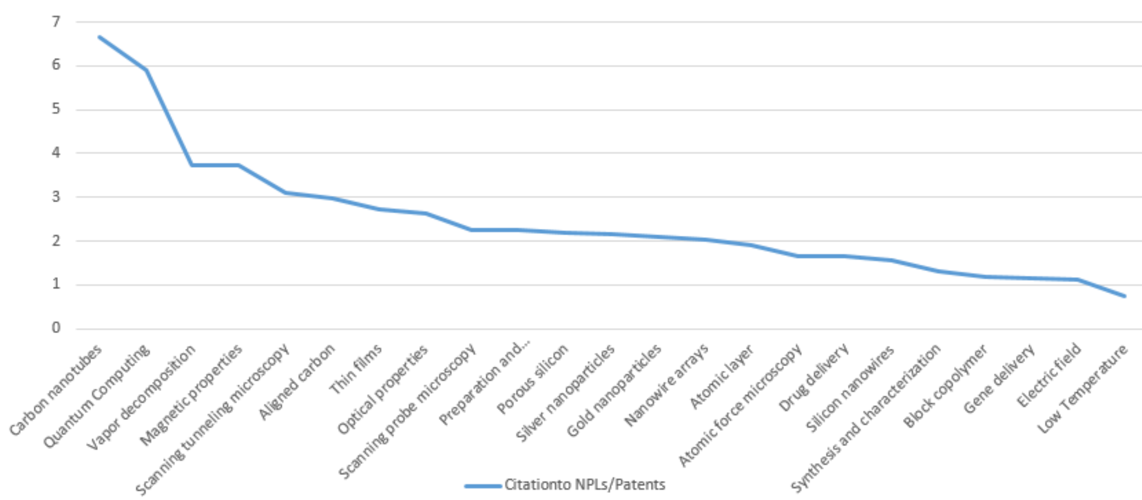
Looking at Table 6, almost all of our top citing patents with highest out-degree value are from period of 2003 to 2007 time slice of our research time line. Another interesting point in Table 6 is related to "assignee" column as we see 7 out of 10 leading citing patents belong to the same assignee "D-Wave Systems, Inc.". The most important nodes in connecting technological and scientific clusters in our patent/article network also belong to "carbon nanotubes" cluster as the main article/patent cluster. D-Wave Systems, Inc. is a quantum computing company, based in Burnaby, British Columbia, Canada. D-Wave's high number of scientific references is justified as we found they employed several key members of the scientific community on a permanent or contract basis. This company is contributing to development of scientific articles as well as new patents as Cho (2014) described one of recent D-Wave studies as the most thorough and precise study that has been done on the performance of the quantum machines.

Similar to our analysis on scientific clusters citations per article average, here we calculated average of citations per patent for summations of out-degree value on each of technological clusters as in Table 7. We used this value as a metric of the connectivity of each technological cluster to scientific domain in NPLs set. It means the higher value of citations per pattern a technological cluster have, the more it is influenced by scientific domains. Looking at Figure 20, we can see stronger citation rate of clusters such as "carbon nanotubes" and "quantum computing" to scientific domains comparing to other domains. The relatively high number of citations to the scientific literature in technological clusters with highest out-degree values in patent-article citation network shows that science seems to drive much more knowledge growth in these fields field than in other clusters with more patent citations. This means these technological clusters with less NPL citations have been developed and influenced by technology domains.

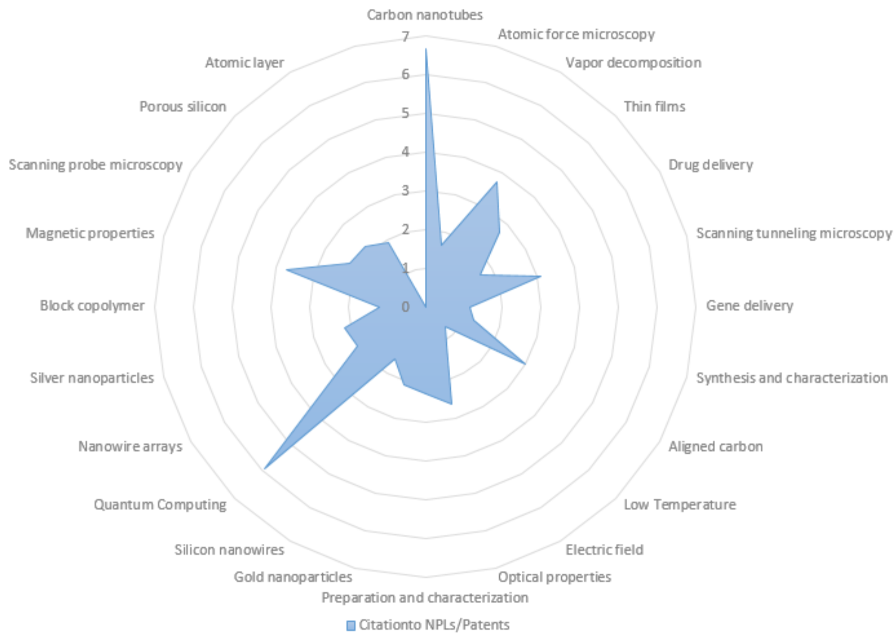
Table 7: Average of citation per patents in technological clusters

Technological Clusters	**Patent**	**Citations (Out-degree)**	**Citation to NPLs/Patents**
Carbon nanotubes	648	4320	6.66
Quantum computing	104	615	5.91
Vapor decomposition	573	2145	3.74
Magnetic properties	92	343	3.72
Scanning tunneling microscopy	270	840	3.11
Aligned carbon	144	430	2.98
Thin films	480	1302	2.71
Optical properties	127	332	2.61

Scanning probe microscopy	90	203	2.25
Preparation and characterization	120	270	2.25
Porous silicon	88	193	2.19
Silver nanoparticles	94	203	2.15
Gold nanoparticles	119	250	2.1
Nanowire arrays	98	198	2.02
Atomic layer	85	162	1.9
Atomic force microscopy	574	950	1.65
Drug delivery	410	678	1.65
Silicon nanowires	115	179	1.55
Synthesis and characterization	160	210	1.31
Block copolymer	94	111	1.18
Gene delivery	243	279	1.14
Electric field	130	147	1.13
Low Temperature	138	103	0.74



(a)



(b)

Figure 20: (a) Citations per patent average of technological clusters (b) Radar chart of citations per patent average of technological clusters

Comparing results of this section to similar research works, as mentioned in our literature review, one of our main references for the study of patent and NPLs was the research done by Karvonen and Kässi in 2013 who studied the impact of science on technology applications in converging technological environments. Similar to their method, we also categorized our citations into technological and scientific categories to see how different patent clusters are developing based on ideas' flow from scientific clusters. According to their results, focusing on paper industry, they found technological convergence have not meant converging knowledge basis in scientific fields. But here we found influenced domains from scientific fields such as "carbon nanotubes", "quantum computing" and "magnetic properties" with the highest rate of citations to scientific articles.

3.2.3 Scientific and Technological Clusters' Evolution

We studied the growth of each scientific cluster in terms of the number of articles published in different years over our time line. Here, in this section, we will present trends of our top 5 clusters according one of our thesis objectives in investigation of leading scientific and technological clusters. Looking at the trends we can see that "carbon nanotubes" has higher number of nodes in comparison to other top clusters as discussed before regarding growth of top scientific domains

(Figure 12), we saw some scientific domains such as "force microscopy" had more publications as it shows scientist's interest in this scientific field back to earlier years while the growth of articles in this field went on in lower rate comparing to "carbon nanotubes" and "vapor decompositions" on which scientists started working later on but a high number of publications between 2001 and 2005 made out of them important fields.

As we know, trend charts are used to show trends in data over time and increase understanding of the real performance of a process, particularly with regard to an established target or goal. Trend lines are used as one of analysis tools in our research since they help us to track growth of different scientific and technological clusters over time, while stacked graphs depict items stacked one on top of the other column or side-by-side (bar), differentiated by colored bars or strips. A stacked graph is also useful for looking at changes in, for example, expenditures added up over time, across several products or services. The graph integrates different data sets to create a richer picture of (the sum of) changes (Kriebel, 2012).
Looking at plots in Figures 21 and 22, they show the distribution of articles in different years in regard to scientific clusters that they belong to. Once again as other plots show, we can see interest in "carbon nanotubes" and "atomic force microscopy" cluster after 2005 among scientific domains, we can also see that scientists' interest in these scientific domains have increased since new ways of producing single-walled carbon nanotubes such as induction thermal plasma method were discovered (implemented in 2005 by groups from the University of Sherbrooke and the National Research Council of Canada).

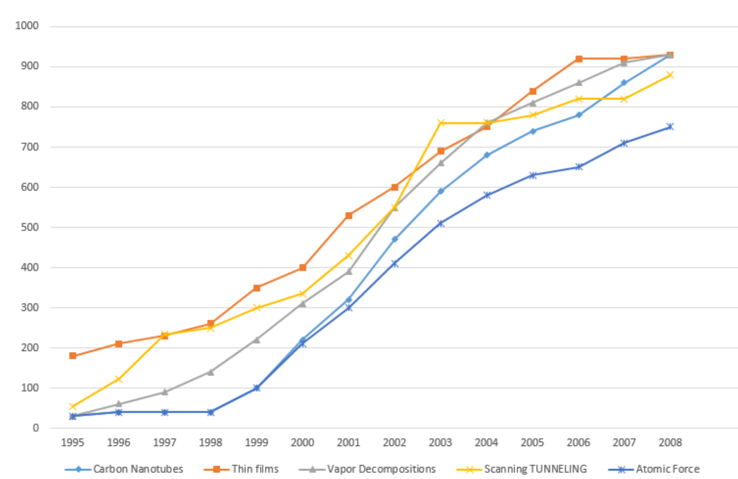


Figure 21: Top 5 cluster trends in NPLs

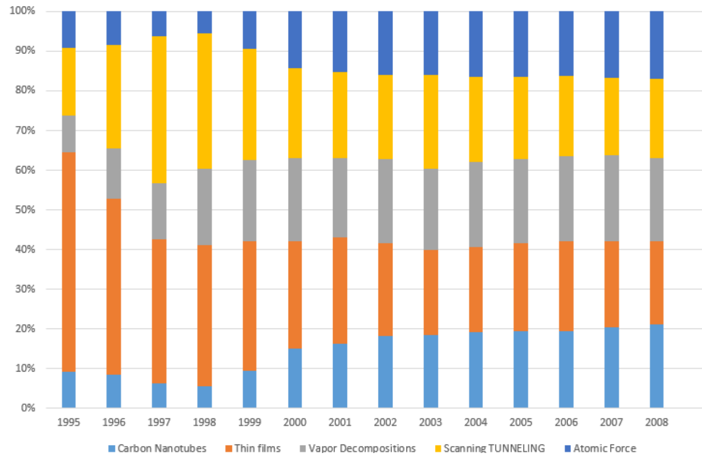


Figure 22: Stocked column 100% chart, top 5 scientific clusters over time in NPLs

Innovation timeline in Figure 23 shows how different scientific clusters are distributed in our timeline. We can see almost all of top clusters are covering all the way of our research timeline which is expected as these scientific domains have the highest numbers of articles in comparison with others. One of the interesting results of innovation timelines is the fact that we have clusters which came into scientist's interest only in 1996 or 1997 and now we have them between our top 15 nanotechnology scientific domains as scientists showed a lot attention and interest in them. This is verified by looking at the high rising trend of the number of publications in these specific clusters such as silver and gold nanoparticles. Looking at the history of nanotechnology fields, nanoparticles were more of scientist's interest after discovery of carbon nanotubes in 1991. In 1995, Nano-imprinting by S. Y. Chou (University of Minnesota, USA) and in 1996, Nano sheets discovery by T. Sasaki in National Institute for Research in Inorganic Materials of Japan were among the remarkable research results related to the nanoparticles research development (Horikoshi and Serpone, 2013).

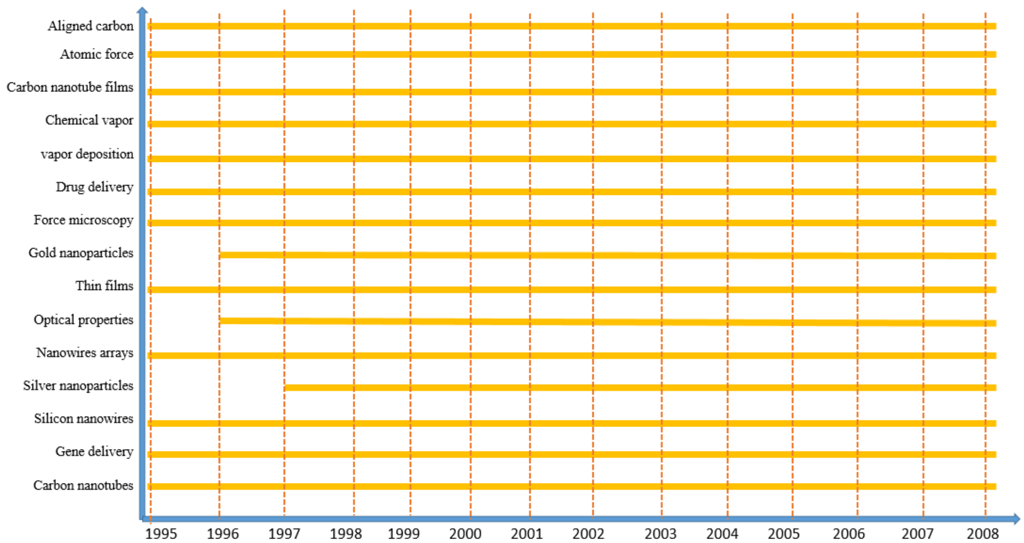


Figure 23: Top leading clusters' innovation time line

Similar as top scientific clusters, in Figures 24 and 25 we can see top 5 technological clusters showing the number of patents in each cluster. Regarding the trend lines we can see some technological clusters such as "carbon nanotube films" have drastic increase in number of patents after year 2000 which can be related to the increase in number of research articles in these technological fields. The more research studies is published in each of these fields that can lead to higher number of patents as we discussed in literature review of the interaction between scientific and technological fields. On the other hand we can see clusters such as "gene delivery" which although we have higher number of records in the earlier years of our research timeline, trend goes on with only a slowly increasing rate.

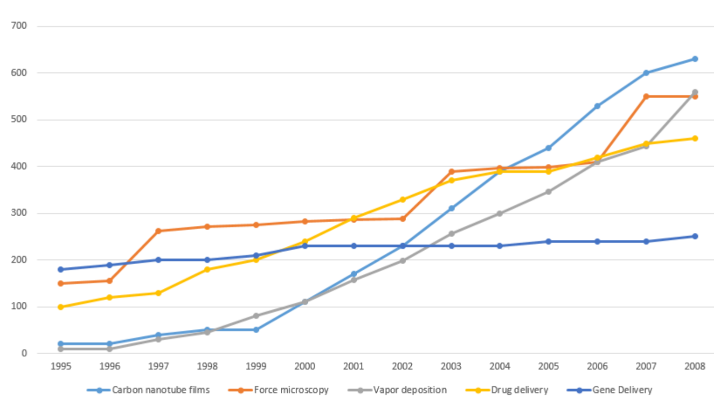


Figure 24: Top 5 cluster trends in patents (Cumulative)

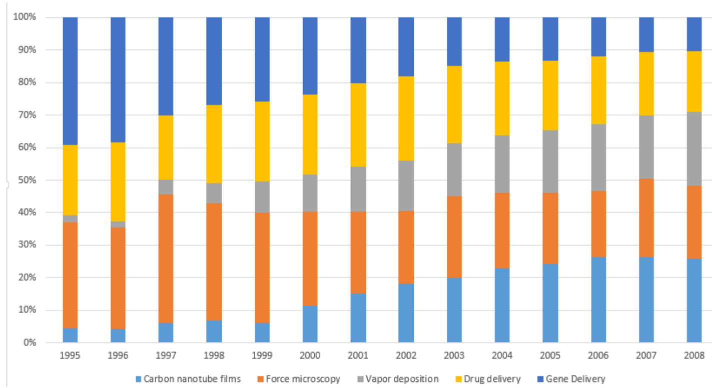


Figure 25: Stocked column 100% chart, top 5 technological clusters over time in patents (Cumulative)

Looking at Figure 26, we visualized top citing clusters patent and top cited article cluster mapping in large citation network of articles and patents. Technological clusters are placed on top of Figure 26 as we showed them by the "Tech-" prefix on labels, while scientific clusters can be seen with the largest cluster of "carbon nanotubes" in bottom. As we mentioned in interpretation of top cluster mappings, clusters with larger overlap areas are more related to each other and can have more interactions in terms of common article or patents. Here we can see overlaps of "carbon nanotubes" with almost all other scientific domains, while it is also strongly correlated with "vapor decomposition" in technological domains. The separation of technological and scientific domains is also justifiable as they are no records in common between articles and patents data.

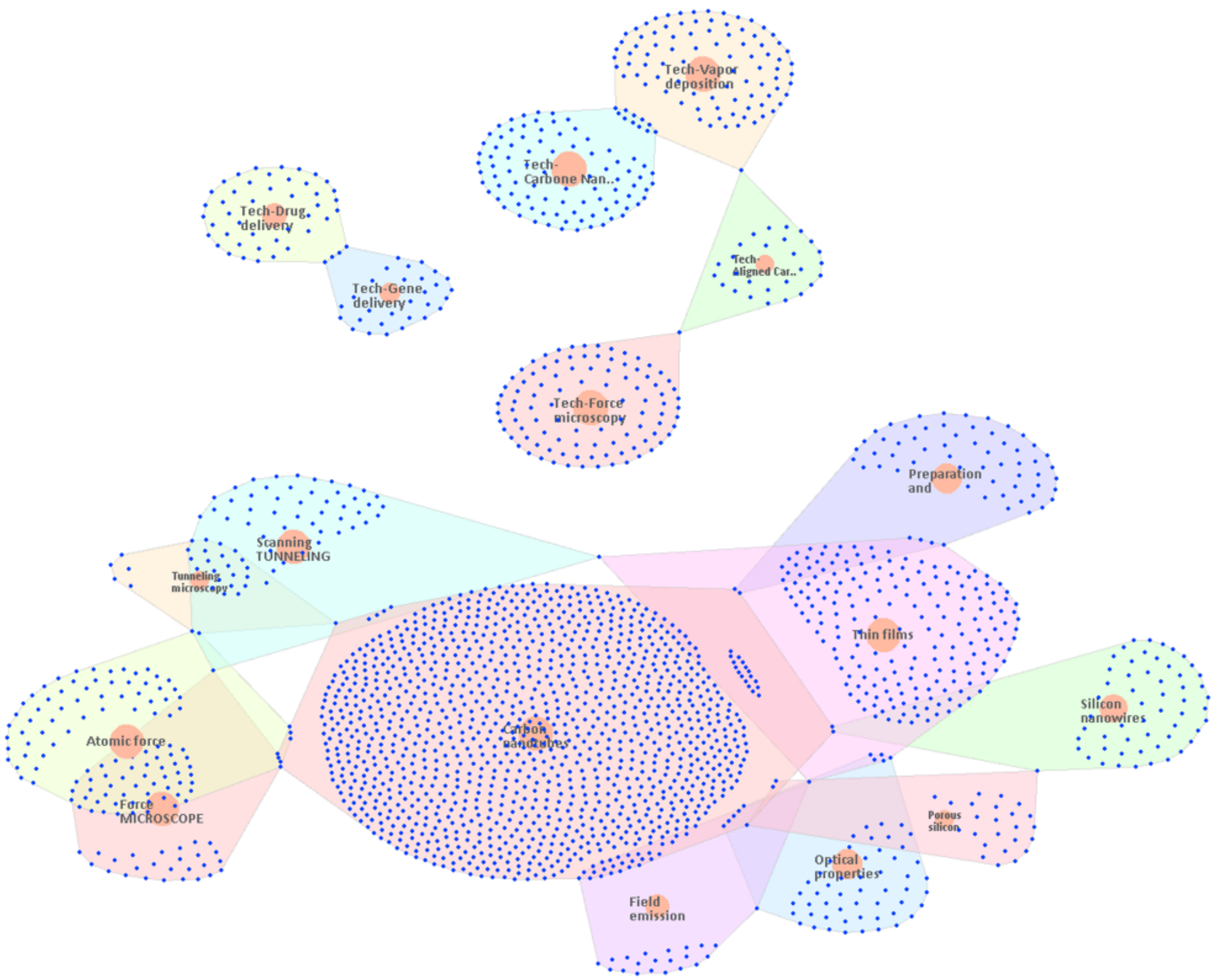


Figure 26: Top technological and scientific clusters in patent-article network

In Figure 27 of our article-patent network, we can see different scientific clusters labels on network while clusters are colored differently for a more clear view. Labels next to each node cluster show the name of each scientific domain and nodes are sized based on in-degree value. In means the bigger a node is the more cited it is by other nodes. "Carbon nanotubes" with the most number of nodes has most interconnected edges with other clusters around while in includes the most cited NPL and article nodes as they can be found by their size. The wide range of "carbon nanotubes" applications in technology is also confirmed regarding our literature review as Sanderson (2006) states the strength and flexibility of carbon nanotubes makes them of potential use in controlling other nanoscale structures, while they will have an important role in nanotechnology engineering.

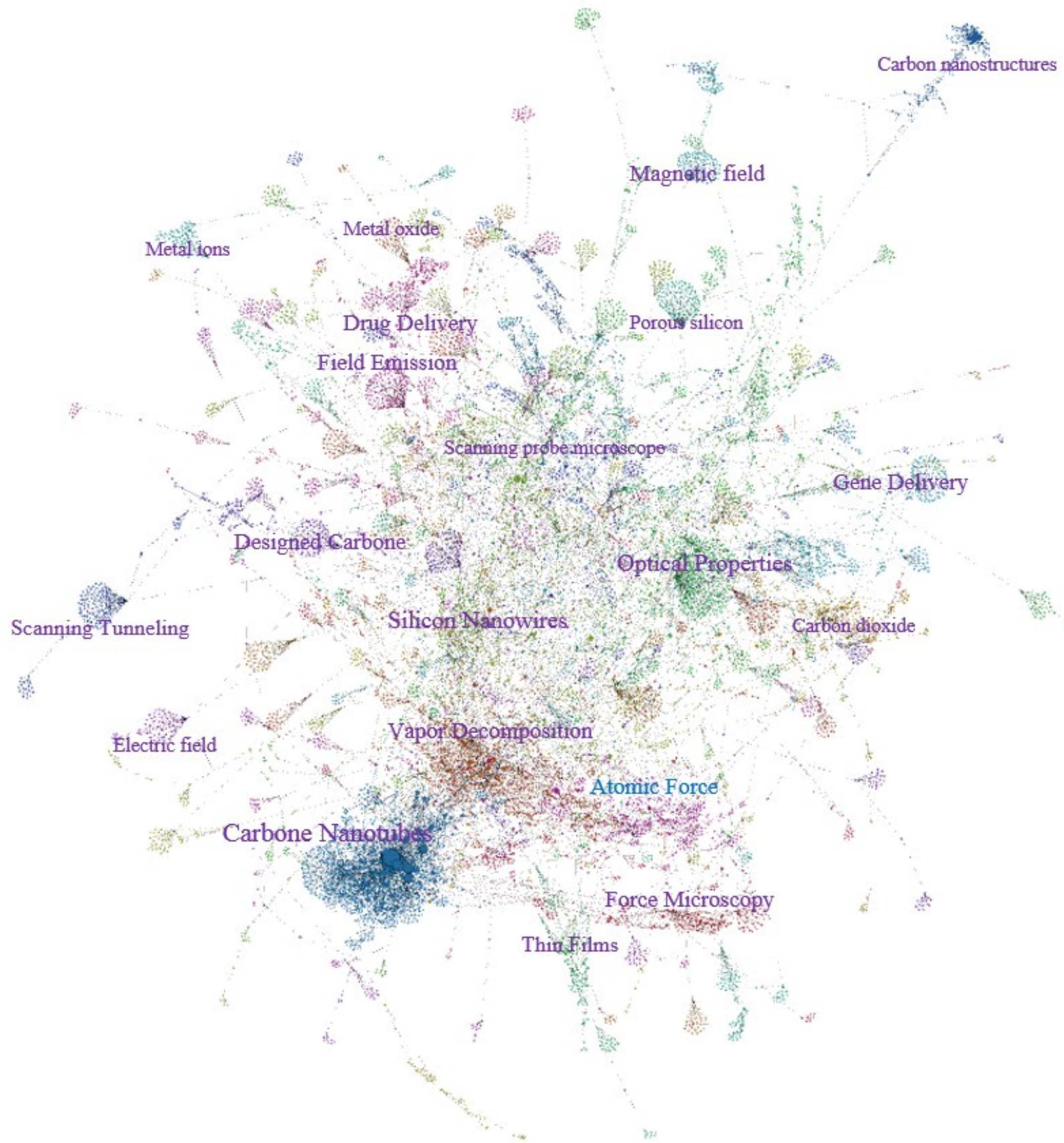


Figure 27: Scientific clusters mapping over patent-article citation network

3.3 Institution and journals contributions

In the next two sub-sections we are going to discuss top journals' and institutions' contribution to development of scientific and technological domains in Canadian nanotechnology.

3.3.1 Institutions' contribution to development of scientific and technological fields

Regarding the data fields provided in our data-set, in this section we did a high level correlation analysis of connections between journals and institutions in Article-NPL sub-network.

There are many research studies on collaborative research between institutions. Some of them such as Melin and Persson (1996), used co-authorship network of authors to study collaboration between different research institutions and investigated how collaboration can be refined and used in study of collaboration on bibliometric data. Owen-Smith and Powell (2008) studied collaboration between institutions in biotechnology focusing on learning from collaborations and how the collaborate happens. Ynalvez and Shrum (2011), investigated the scientific collaboration between institutions to find if it is associated with increased publication productivity. Their findings suggest a new way of modeling publication productivity, with implications for science and innovation policy in both the developed and the developing world.

In this research as we were studying citation network, and since we got institution information of the first author of each article, we built the citation network of Canadian institutions based on their nanotechnology related articles. Looking at this citation network, we can see how different institutions are related to each other through citations. According to Figure 28, It is obvious that University of Toronto is one of the most important nodes of our citation network with a significant share of articles and important In/Out links to other nodes of network. We can see the strongest citations of "University of Toronto" to "University of Alberta" and "University of British Columbia". This shows highest in-degree values between different institutions in our network are among first authors of mentioned universities. "University of Toronto" has the most cited articles in nanotechnology as it has the highest value of in-degree among all nodes. This helps us better understand the contribution of different Canadian institutions in nanotechnology related articles and how they are related in terms of citation relations. Also by looking at Table 8 we can see that "University of Toronto" has the highest number of articles and in-degree citations among Canadian institutions in nanotechnology related fields, which can justify the higher number of in-degree citations to the articles published by first authors from this university. Looking at Figure 28, considering the high number of nodes in citation table, we classified them into two groups of top highly cited institutions highlighted by orange circles and other institutions with lower rate of in-degree values with blue circle, while the size of nodes shows the in-degree value. Size of nodes is considerably bigger for "University of Toronto" and "National Research Council of Canada" as they have much more in-degree citations comparing to other institutions.

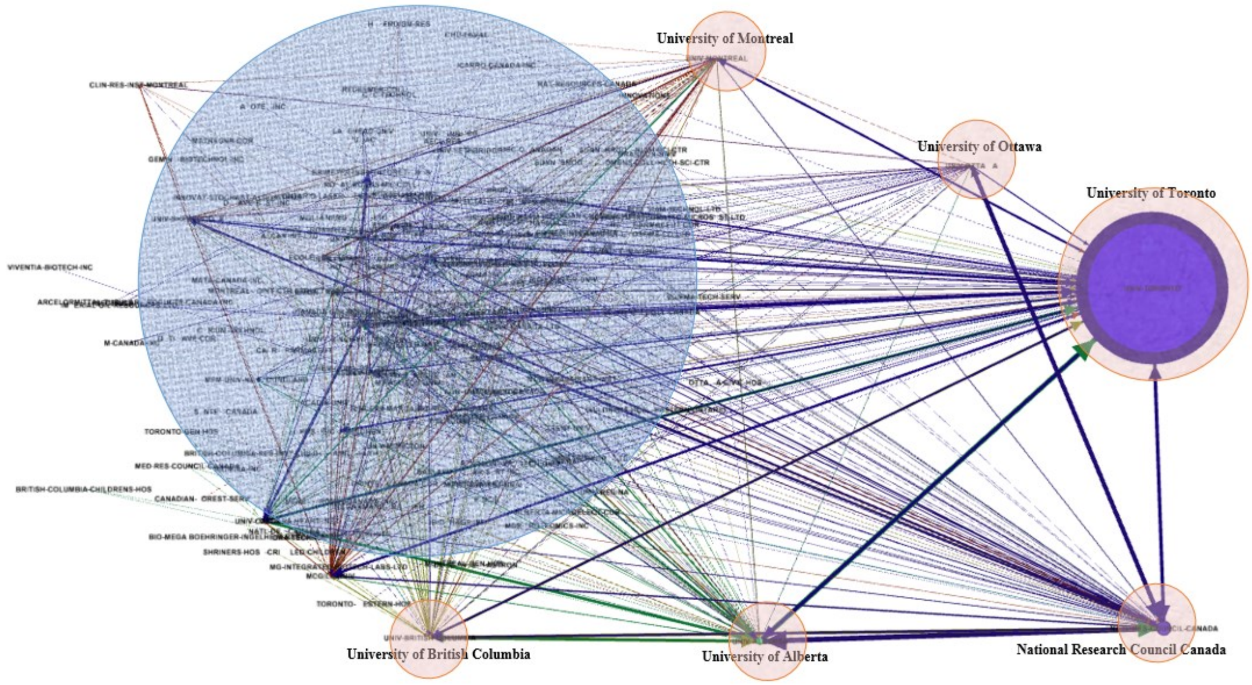


Figure 28: Institutions contributions citations in article citation network

Based on Thomson Reuters (2008), ranking institutions in terms of paper counts helps to compare the productivity and volume of research output among various institutions. To have a better view of Canadian institutions’ share in publishing articles in nanotechnology related fields we can take a look at Table 8 and the plot of Figure 29. It is clear that among top 40 institutions which publish nanotechnology related articles “University of Toronto” has the highest share of articles in our data-set. This is an expected results as University of Toronto has established Canada's first center for nanotechnology research, formed in September 1997 under the name The Energenius Centre for Advanced Nanotechnology (ECAN). Strong industrial support, a team of world-leading research scientists and state-of-the-art tools place CAN at the forefront for developing the key enabling technologies, nano-electronic and nano-photonic applications (Ruda, H., 2008).

Table 8: Top 40 Canadian institutions contribution to nanotechnology scientific domains

Rank	Institution	Citations/Article
1	UNIV-TORONTO	35.47
2	NATL-RES-COUNCIL-CANADA	26.34
3	UNIV-ALBERTA	35.89
4	UNIV-BRITISH-COLUMBIA	36.63
5	MCGILL-UNIV	34.66
6	UNIV-WATERLOO	28.49

7	MCMASTER-UNIV	31.91
8	UNIV-WESTERN-ONTARIO	36.68
9	UNIV-MONTREAL	38.12
10	UNIV-CALGARY	33.51
11	UNIV-OTTAWA	37.89
12	SIMON-FRASER-UNIV	27.68
13	QUEENS-UNIV	32.21
14	UNIV-LAVAL	34.41
15	ECOLE-POLYTECH	23.93
16	UNIV-SHERBROOKE	29.01
17	UNIV-SASKATCHEWAN	33.52
18	UNIV-GUELPH	39.57
19	UNIV-MANITOBA	37.32
20	DALHOUSIE-UNIV	33.4

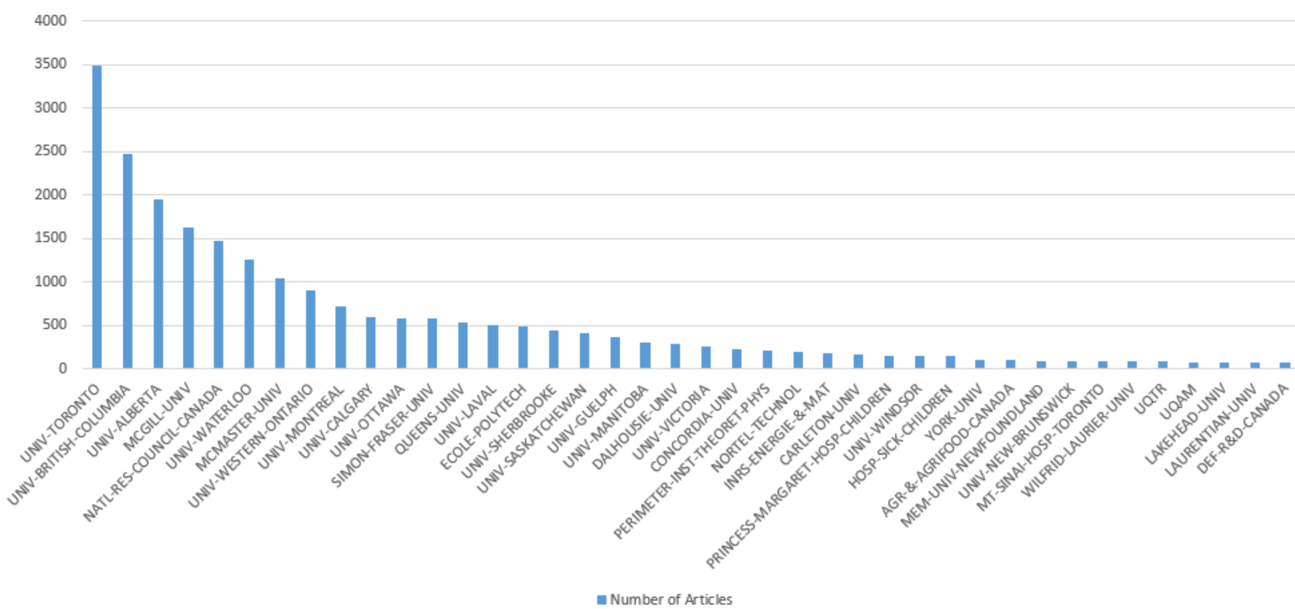


Figure 29: Top 40 Canadian institutions contribution to nanotechnology scientific domains

In the analysis of Table 9 and Figure 30, we showed the trend of citations per article for each of the institutions above. Looking at the trend of citations per article among all the top 40 institutions, we found that institutions such as “University of Toronto”, “University of Alberta” and “University of British Columbia” with higher number of papers are almost in the same range of citations per article too. On the other hand, it was interesting that top institutions with highest citations per article value were among medical institutions of “Mount Sinai Hospital, Toronto”, “Hospital of Sick Children” and “Princess Margaret Hospital of Children”. From this, we can conclude higher citations per article on nanotechnology related articles in medical and health domains. Comparing these results with research literature, the emerging landscape of nano-medicine is also confirmed by Wagner and his colleagues (2006), as they did a global survey of companies pursuing nano-medicine, and confirmed that nanotechnology is taking root in drug and medical domains.

Table 9: In-degree citation per article for top 40 Canadian institutions

Rank	Institution	Citations/Article
1	MT-SINAI-HOSP-TORONTO	49.56
2	HOSP-SICK-CHILDREN	48.24
3	PRINCESS-MARGARET-HOSP-CHILDREN	43.78
4	UNIV-GUELPH	39.57
5	UNIV-MONTREAL	38.12
6	UNIV-OTTAWA	37.89
7	UNIV-MANITOBA	37.32
8	UNIV-WESTERN-ONTARIO	36.68
9	UNIV-BRITISH-COLUMBIA	36.63
10	UNIV-ALBERTA	35.89
11	UNIV-TORONTO	35.47
12	UQAM	35.09
13	INRS-ENERGIE-&-MAT	34.8
14	MCGILL-UNIV	34.66
15	UNIV-LAVAL	34.41
16	YORK-UNIV	34.21
17	AGR-&-AGRIFOOD-CANADA	33.98
18	LAKEHEAD-UNIV	33.6
19	UNIV-SASKATCHEWAN	33.52
20	UNIV-CALGARY	33.51

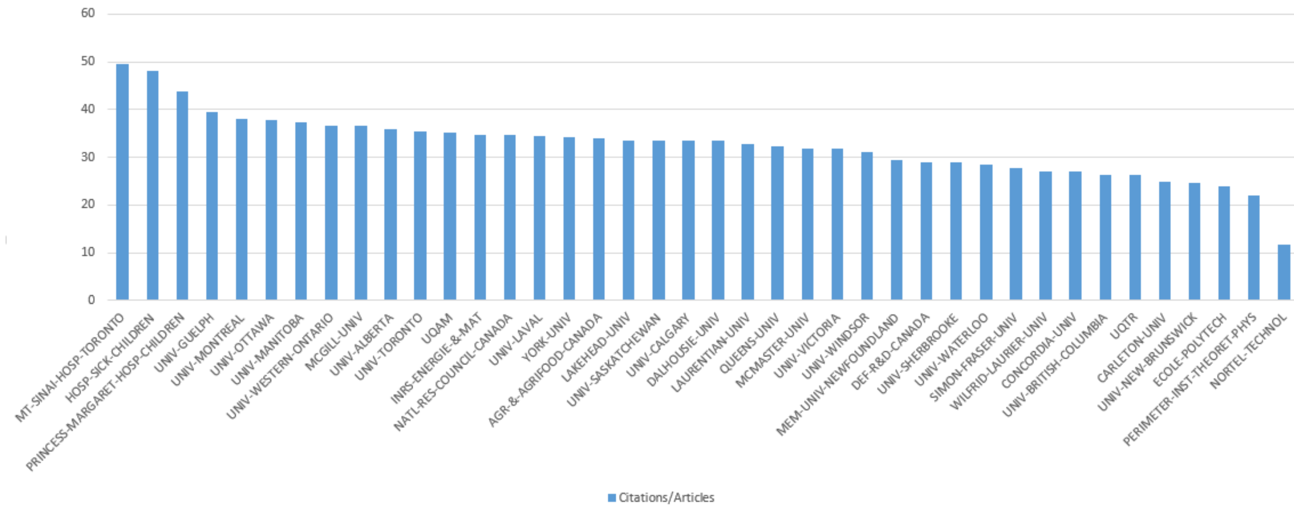


Figure 30: In-degree citation per article for top 40 Canadian institutions

Discussing citations per article metric as a quantitative measure of institutions' impact, we extracted trend of this value while sorting institutions based on summation of their in-degree citations. According to Thomson Reuters, (2008), Citations per paper (sometimes called "impact") is computed by dividing the sum of citations to some set of papers for a defined time period by the number of papers (paper count). The citations per paper score is an attempt to weight impact in respect to output, since a greater number of publications tends to produce a greater number of citations. Results of the chart in Figure 31, show a rising trend for citations/articles, which shows, by increasing the number of citations to articles belong to research papers from a sample institution, the impact they have on development of scientific fields will also increase.

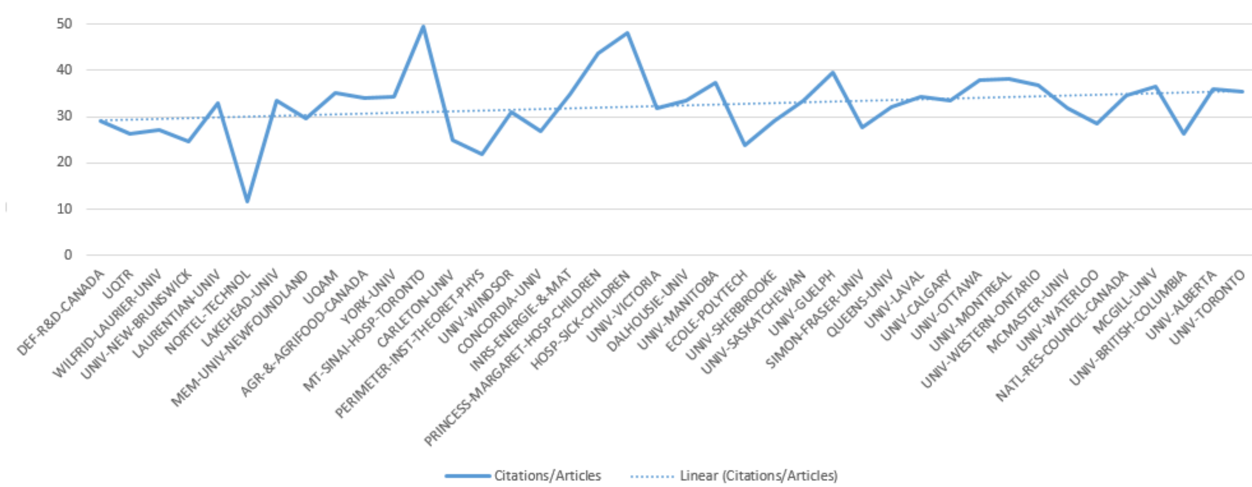


Figure 31: Institutions sorted based on citations count, trend shows citations per article

To investigate the contribution of Canadian institutions to technological fields, we present Figure 32 which is extracted based on citations from patents to NPLs. As the chart shows, “University of Toronto” has a considerably higher rate of in-degree citations from patents side, which shows more patents are influenced by the publications of research works in this university. These results were expected according to high rate of “University of Toronto” impact on scientific fields and their recent efforts in development of nanotechnology related fields.

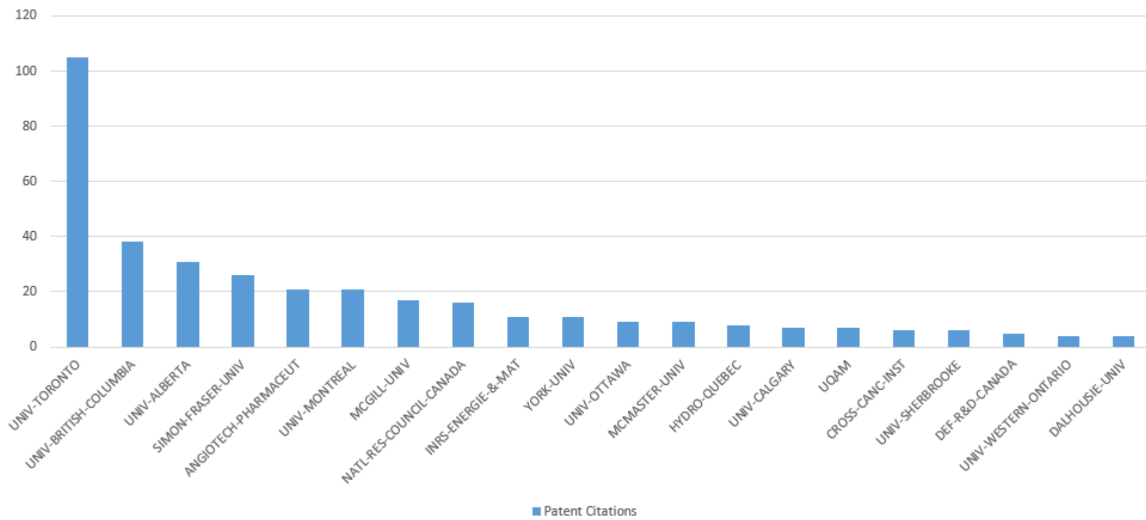


Figure 32: Institutions sorted based citations count to patents

Looking at Table 10 and Figure 33, we extracted citations value as a measure of institutions impact on technological clusters. Results show highest rate of citations from patents per articles for “University of Toronto” while “UQAM” is in the next rank. The high impact rate for “UQAM” and “Hydro Quebec” among Canadian institutions can be related to their relation to industry and also some highly cited articles belonging to these institutions which caused a significant increase in the citations/article value for them.

Table 10: Canadian institutions patent citation per article

Rank	institution	Citations/Article
1	UNIV-TORONTO	0.1009
2	UQAM	0.0853
3	CROSS-CANC-INST	0.08
4	HYDRO-QUEBEC	0.0727
5	INRS-ENERGIE-&-MAT	0.0588
6	SIMON-FRASER-UNIV	0.0448
7	DEF-R&D-CANADA	0.0406

8	ANGIOTECH-PHARMACEUT	0.0308
9	YORK-UNIV	0.0301
10	UNIV-MONTREAL	0.0291
11	UNIV-BRITISH-COLUMBIA	0.0233
12	UNIV-ALBERTA	0.0158
13	UNIV-OTTAWA	0.0154
14	DALHOUSIE-UNIV	0.0139
15	UNIV-SHERBROOKE	0.0134
16	UNIV-CALGARY	0.0117
17	MCGILL-UNIV	0.0115
18	MCMASTER-UNIV	0.0085
19	NATL-RES-COUNCIL-CANADA	0.0064
20	UNIV-WESTERN-ONTARIO	0.0044

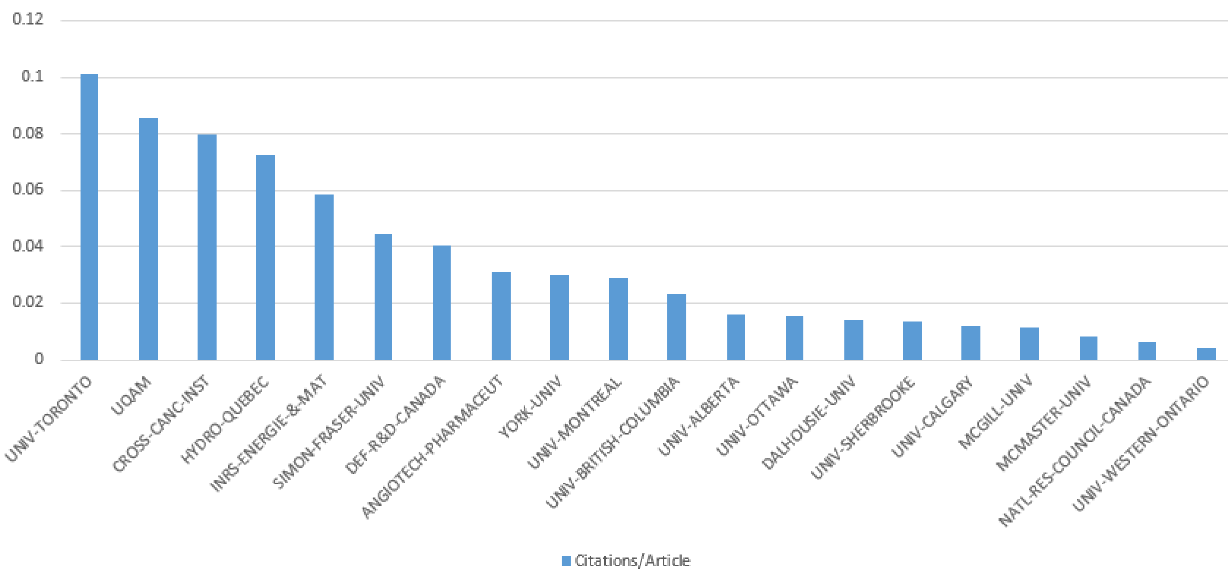


Figure 33: Canadian institutions citation to patents per article

3.3.2 Journals' contribution to development of scientific and technological fields

Considering Leydesdorff's (2006) study on journals' interdeciplinarity using degree centrality parameters, and also regarding one of our objectives, i.e. to investigate contribution of scientific journals in development of nanotechnology domains, we studied the trends of our most cited journals in NPLs set. Most cited journals were of our interest as they were the most referenced journals by patents in different nanotechnology domains. This means that more technological domains are connected to scientific clusters using NPLs published by these journals and this shows their significance for our research.

In addition to centrality measures such as in-degree, we also considered impact factor of journals as a measure reflecting the average number of citations to recent articles published in that journal. Impact factor is used in variety of research studies besides citation count values such as study by McVeigh (2004) on analysis of impact factors and citation patterns. Considering the method of McVeigh in using citation characteristics of journals and impact factor in study of citation patterns and also Leydesdorff (2009) in study of journals interdisciplinary using degree centrality parameters, in the following part of thesis we are going to have a look on trends of different citation characteristics of top cited journals in NPLs set.

Similar to what we discussed regarding scientific clusters, here in Figures 34 and 35, we see how top journals such as "Nano Letters", "Physics Review B" and "Applied Physics Letters" play key role in publishing great number of articles in Canadian nanotechnology, especially in recent years. Among all, "Nano Letters" journal's trend seems more interesting as its share of publications in nanotechnology has a higher growth rate in comparison to other top journals years after 2001, and made this journal became one of our top publishing journals of our research time line.

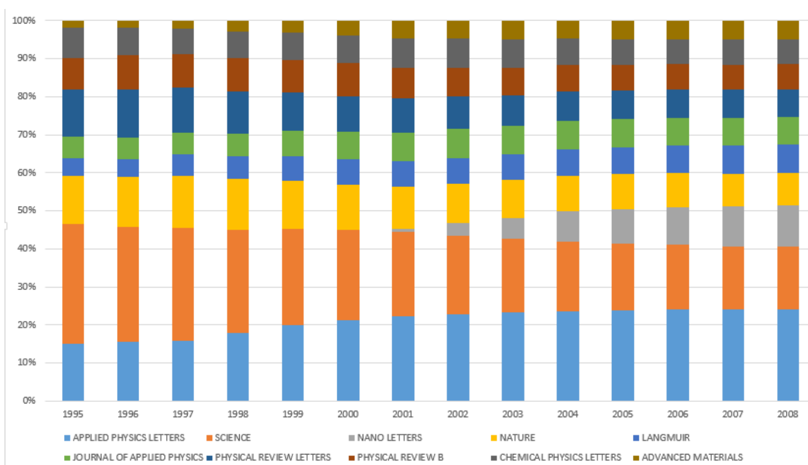


Figure 34: Stacked column 100% chart, scientific journals on research timeline

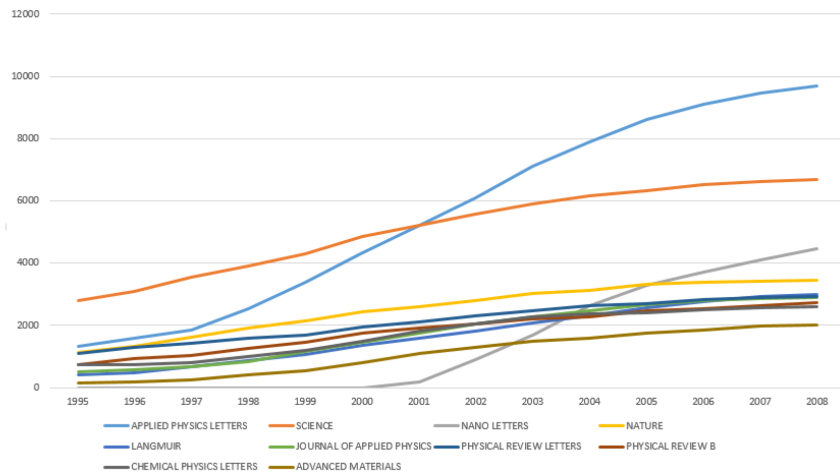


Figure 35: Trend of number of nanotechnology articles in scientific journals on research timeline

Regarding Table 11 and Figure 36, we can see impact factor trend line of top cited journals of NPLs. Journals are sorted based on their in-degree value in ascending order, and the linear trend line added to the chart to show impact factor's trend. Purpose of this chart in Figure 36 is to confirm the validity of our data as in-degree value of journals which shows their number of citations has a positive relation with their impact factor. Figure 36 shows this positive relation as trend of impact factor has a slight ascending trend while we have node in-degree values also sorted ascending. As mentioned, this result shows the significance of top journals in terms of their citation counts in development of scientific domains.

Table 11: Journals' contribution to scientific and technological domains

Journal Name	Number of patent citations	Number of article citations	Number of all citations	Citations/Articles	Impact factor	Scientific clusters covered	Technological clusters covered
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY	265	60866	61131	9.14	12.113	9	8
JOURNAL OF PHYSICAL CHEMISTRY B	142	71291	71433	10.84	3.302	8	6
NANOTECHNOLOGY	342	66291	66633	13.33	3.821	6	7
JOURNAL OF APPLIED PHYSICS	470	157524	157994	17.04	2.183	12	10
MACROMOLECULES	142	83304	83446	21.6	5.8	6	6
PHYSICA E	139	60900	61039	23.48	1.522	4	4
CARBON	365	51512	51877	25.92	6.196	4	3
APPLIED PHYSICS LETTERS	2250	409551	411801	27.49	3.302	14	8
LANGMUIR	390	235001	235391	27.65	4.457	11	9
PHYSICAL REVIEW B	492	386101	386593	28.33	3.736	14	9
CHEMICAL PHYSICS LETTERS	925	87105	88030	32.22	2.337	4	4

MATERIALS LETTERS	137	60006	60143	32.65	2.305	4	2
NANO LETTERS	1839	172019	173858	49.16	13.592	6	5
ADVANCED MATERIALS	435	182935	183370	65.77	17.493	5	6
PHYSICAL REVIEW LETTERS	446	463213	463659	68.50	7.512	9	5
BIOMATERIALS	191	73687	73878	76.47	7.604	2	1
NATURE MATERIALS	342	55046	55388	92.62	36.503	2	2
NATURE	760	259836	260596	183.51	41.456	8	5
SCIENCE	2278	333975	336253	242.78	33.611	9	7
CANCER RESEARCH	217	35229	35446	313.68	9.329	2	2

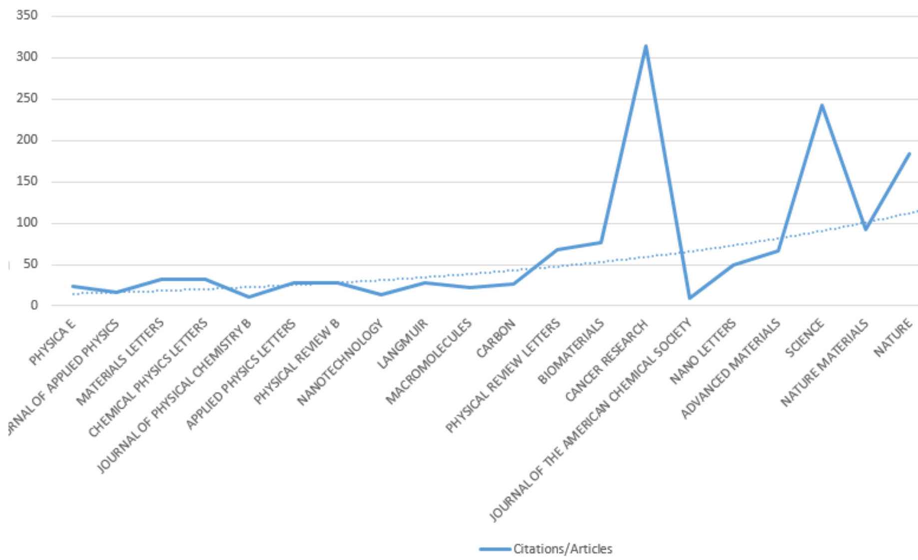


Figure 36: Citations per article in scientific journals

In Figure 37, the scenario is different as we can see trend line of impact factor for top journals, as they are sorted based on their patent citations. This means that in Figure 37, we can see the behavior of journals impact factor regarding their influence on technological domains. Results show a rising trend of impact factor as the patent citation count increases for journals. This positive relation between patent citations count and journal's impact factor is interesting to us since we can see the more articles of a specific journals are cited by patents, the more impact factor the articles of that journal have. In other words, impact factor not only shows the impact of articles on development of scientific domains, it also shows how the articles of a journal have impact on development of technological domains.

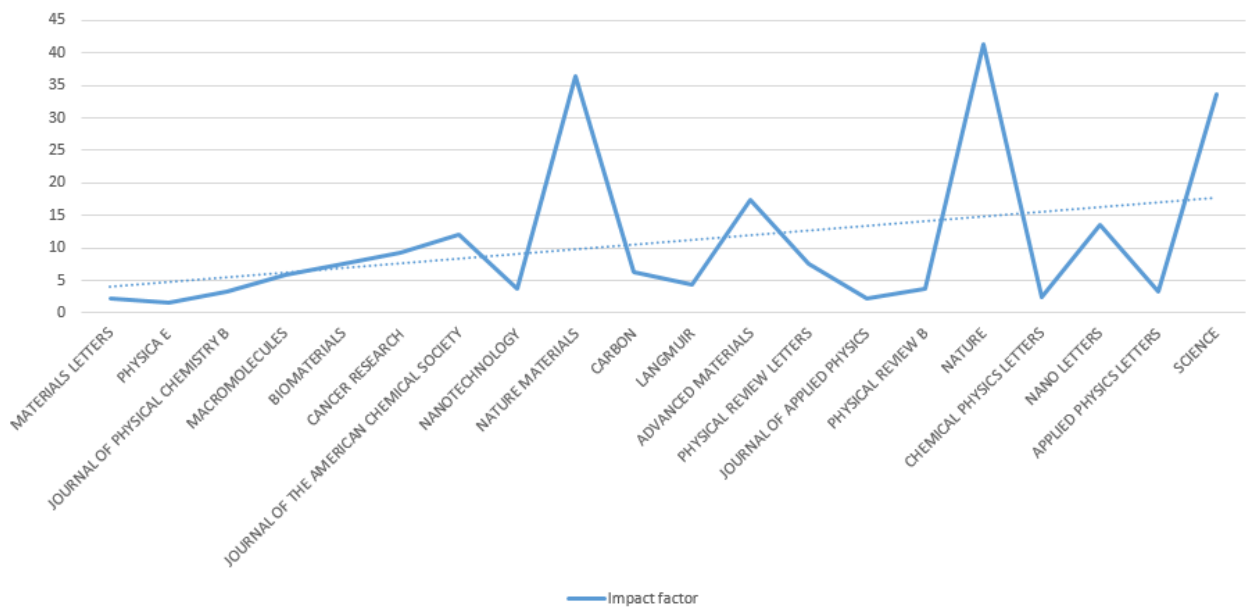


Figure 37: Impact factor trend line considering journals citations from patents

In Figure 38, journals are sorted in descendant order based on in-degree value, or in other words their citation counts, and trend line shows in the number of scientific clusters covered by each of these top journals. We extracted number of covered scientific domains by the information from clustering section and checked to how many different domains each of these highly cited journals have contributed. The results show the increasing trend of contribution to different scientific domains of journals as their in-degree value increases. It is interesting to us as we know the more a journal is publishing articles from different scientific domains, the more multidisciplinary it is. Here we can see in-degree value of a node in the citation network has a positive relation with the number of scientific domains it contributes to. These results are important as they can help us to investigate how topological network parameters values such as in-degree can indicate the multidisciplinary of journals. According to our literature review, Leydesdorff (2006) investigated the interdisciplinary of journals in relation to other centrality degrees such as betweenness centrality, while Igami and Saka (2007) investigated the contribution of articles in development of other scientific fields through citation network of articles from different scientific domains. Comparing these studies with our research, in our analysis we are using in-degree value of nodes as the measure of their contribution to scientific journals and we are focused on nanotechnology sub-fields.

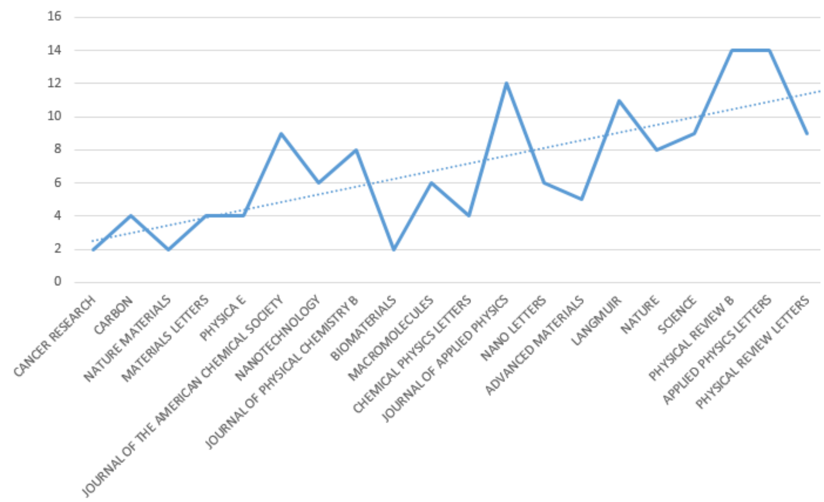


Figure 38: Number of scientific clusters covered by each scientific journal

In Figure 39, we saw the journals are sorted based on the number of in-degree value, or in other words their citation counts. But here in Figure 39, trend line shows number of technological clusters covered by each of these top journals. This chart sheds light on one of our objectives in regards to investigation of journals' contribution to development of technological clusters in our citation network. This result shows the increasing trend of journals' contribution to different technological domains as in-degree value of journals increases.

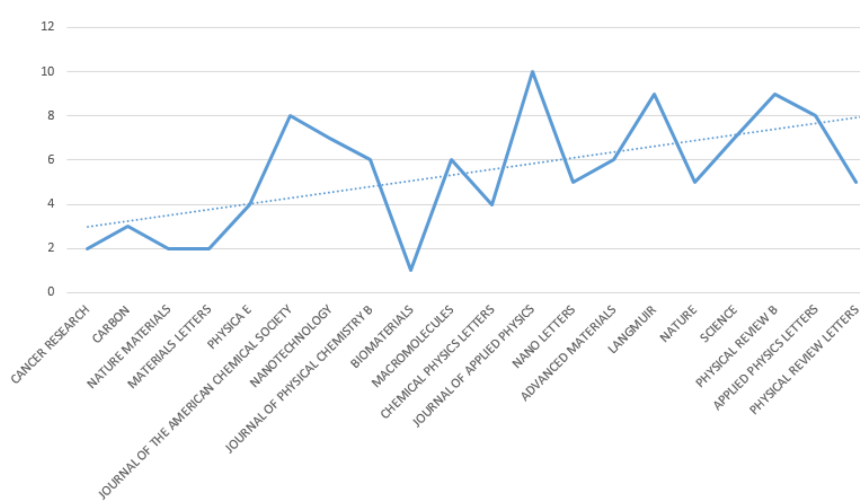


Figure 39: Number of technological clusters covered by each of journals

Comparing the results of analyses above with related research works before, one of the main differences of our study with previous studies is our data as we used more recent data of nanotechnology articles and patents, while similar studies such as Leydesdorff (2009) used older data of communication and McVeigh (2004) was more focused on open access journals in her research. Our results are also different as we investigated the contribution of top cited journals in development of technological domains by study of the number of different technological domain each journal is connected.

3.4 Chapter Summary

In this chapter we discussed our research analysis steps such as cluster analysis, network topological analysis, data trend analysis, and correlation analysis of leading patent/article nodes as well as top main scientific and technological clusters. In the next chapter we are going to conclude the thesis and discuss research limitations and suggestions for future studies.

4 Conclusions, Limitations and Future Studies

Introduction

In this chapter we will conclude the thesis and will discuss research limitations and future studies on this topic. To confirm the originality of our research, we reviewed our literature while covering wide variety of domains such as complex citation networks, social network analysis, text mining clustering methods and research studies on interactions between scientific and technological domains. We found although there are previous research papers which investigated patent and article citation networks, this study investigates the citation network of both article and patents in one large network over a longer research time line comparing to similar research works. Moreover, according to our literature review, this study is the first to examine the flow of ideas from scientific to technological fields over the patent-article citation network by extracting leading articles and patents which play a crucial role in keeping this knowledge flow alive in Canadian nanotechnology industry.

4.1 Conclusions

This research explores interaction between scientific and technological domains by studying citation networks of articles and patents in the Canadian nanotechnology. As one of the main research objectives, we investigated the relationship between network topological measures and article/patent nodes in order to find leading article and patents within the network which play critical role in connecting scientific and technological nanotechnology clusters.

The first objective of this research was to investigate most important NPLs and to examine their significant gate-keeper role in connecting scientific and technological domains. By comparing NPL citations to sets of articles and patents, results show higher range of NPLs' contribution to development of scientific fields comparing to technological domains in Canadian nanotechnology. We investigated top cited NPLs which technological domains were most influenced by, as well as most citing NPLs which established a strong connection between scientific and technological clusters of our patent-article network. Using average of citations per article metric, we calculated which different technological domains influenced more by scientific NPLs and also the impact of

NPLs on development of various scientific domains. Results showed, "scanning probe microscopy" as the scientific domain with highest value of citations per article average, which shows high rate of this cluster's impact on technological domains as well as the significance of highly cited articles in this field. Considering nature of "scanning probe microscopy" field, we can realize significant role of articles in this cluster in development of patents and other articles as we have high number of articles in other clusters such as "atomic force microscopy" and "scanning tunneling microscopy" citing to articles in this field. On the other hand, among technological clusters, we saw higher citation rate of clusters such as "carbon nanotubes" and "quantum computing" to scientific domains comparing to other technological domains. The relatively high number of citations from these technological clusters to the scientific ones shows that science seems to drive much more knowledge from these fields other than clusters with more citations to technological clusters.

Regarding our next objective, we extracted data cluster maps and investigated growth of scientific and technological clusters in our research time line. Results showed the earlier growth of some core scientific domains such as "carbon nanotubes" and "vapor decompositions" according to the interest scientists were showing to these fields since 1950s. While on the other hand some other top fields such as "scanning tunneling microscopy" got more advanced in recent years as they had most of their growth between 2004 and 2008. Regarding clusters' overlaps, we found "carbon nanotubes" cluster has overlaps with higher number of other clusters. This indicates more articles it has in common with other scientific domains and therefore shows higher level of multidisciplinary of this domain in terms of more relations with other fields. Looking at technological cluster maps, we found "carbon nanotubes" cluster as the cluster with highest number of patents, has biggest overlap area with "vapor decomposition" in technological domains which shows the higher number of patents these two domains have in common and therefore stronger links they have in between. Also looking at scientific cluster correlation maps of our network, "carbon nanotubes" has most interconnected edges with other clusters while it includes the most cited NPL and article nodes. The wide range of "carbon nanotubes" applications in technology is also confirmed by our literature review as Sanderson (2006) states the strength and flexibility of carbon nanotubes highlight their potential use in controlling other nanoscale structures, while they will have an important role in nanotechnology engineering.

Regarding investigation of the correlation between topological network quantitative parameter values such as citations per articles and journals' impact factor as the qualitative parameter, we found citation count value of journals in our citation network has a positive relation with the number of scientific domains it contributes to. In addition, we found the positive relation between patent citations count and journal's impact factor. This is interesting to us since we can see the more articles of a specific journals are cited by patents, the more impact factor the articles of that journal have. In other words, impact factor not only shows the impact of articles on development of scientific domains, it also shows how the articles of a journal have impact on development of technological domains. Regarding the NPL's contribution to development of technological domains, we found a positive relation between NPL journals' citations to technological clusters and the number of technological clusters they cover. Results showed the increasing trend of journals' contribution to different technological domains as citation count value of journals increases. Comparing our results with similar research studies in this field, our results are covering a wider scope as we as we focused our investigation on the contribution of top cited journals in development of technological domains by studying the number of different technological domains each journal is connected to. Regarding the contribution of top cited NPL articles in development of scientific domains, we discovered a positive correlation between, betweenness centrality measure and citation count of articles, which indicates the more an article is cited by patents and other articles, the more influence it has on the transfer of ideas from scientific to technological domains.

Regarding the investigation of top journals and Canadian institutions impact on development of scientific and technological domains, we found "University of Toronto", "University of Alberta" and "University of British Columbia" have higher number of articles in scientific clusters. The results are expected as we already discussed University of Toronto's efforts in development of center for advanced nanotechnology researchers to establish critical mass of principal investigators and facilities to enable them to perform internationally competitive research. In addition, regarding the most cited institutions, we found higher citations per article on nanotechnology related articles in medical and health domains. The emerging landscape of nano-medicine as it is taking root in drug and medical domains, is also confirmed by Wagner and his colleagues (2006), as they did a global survey of companies pursuing Nano-medicine. Regarding the contribution of Canadian institutions in the development of technological domains, we found higher contribution rate of “UQAM” and “Hydro Quebec” among Canadian institutions. This can be related to these institutions’ strong links to industry and also because of some highly cited articles belonging to these institutions which made their citations/article parameter value increase drastically.

4.2 Limitation of Study

This research was exposed to some limitations as we discuss in this section:

One of our main concerns was about data-set completeness and to cover all nanotechnology related articles and patents within our research timeline. Our data gathering and data preprocessing steps were of most time consuming parts of this research as we needed to gather data from our references and remove all noise records. Although we got citation information of articles up to year 2008, but still concerns about completeness of citation data exists.

Although we spent a few months of research work on data preprocessing, we cannot guarantee 100% of data accuracy. Data preprocessing steps were about parsing raw article and NPL strings, tokenizing them, and then after all these data was matched by WOS database to guarantee the right format of records. But still, as we were dealing with huge number of records, still in the last iteration of data clustering we had few numbers of articles in “other languages” or “wrong string” clusters which we needed to discard them. We also missed some data records as we needed to extract data and match it with anther data-set in another format to get full records information.

As the next data related limitation, the largest component of our network was covering 98% of all nodes, which shows high connectivity of nodes in our main network. But still about 2% of nodes were discarded as we needed to just work on all countries network’s huge component.

The standard practice in bibliometric research is to normalize citations in respect to the subject category and publication year. In this research, we used number of citations per article/patent as an attempt to weight impact in respect to output, assuming that a greater number of publications tends to produce a greater number of citations. We did not normalize citation in terms of the subject category as all our data was related to nanotechnology field and we did not have sufficient data to weight each subfield in terms of their scientific output.

One of other limitations we faced, was variety of parameters which can impact a node to be selected as leading network nodes. As we know, there are informal relationships between authors or author inventors which might lead to a future citation relation in science and technology citation networks which we couldn't track in our data-set.

4.3 Future Studies

Future studies of this research are proposed based on the research limitations and also the ideas we got during the time we did this research:

We did our research looking at both network of Canadian and also records from all countries in nanotechnology domains. As we got country and province information for all of our records, it will an interesting idea to do geographical analysis on the patent–citation network and investigate how different Canadian provinces contribute to the development of Canadian nanotechnology. Future research works can also investigate different institutions from different regions interact in knowledge transfer through citation relations in Canadian nanotechnology.

As we were reviewing text mining based methods of data clustering, topical cluster analysis seems to be an interesting idea to work on. We can extract word co-occurrence networks units of science that can be calculated by an analysis of the co-occurrence of words in associated texts. Co-occurrence mappings can get extracted similar to what we did in cluster mappings and the structure and evolution of a research domain can also get analyzed by this method.

Recommender systems based on prediction models have become extremely common in recent years. One another interesting ideas which can followed on this research is to work on a prediction model top investigate the growth and evolution of each scientific/technological cluster in future years. The model can be built based on neural network methods and trained on our research timeline, then results will be useful for research and governmental agencies in assigning resources to each science and technological domains.

In this research, we studied a sample pattern of knowledge flow in our citation network, from science to technology domains. Another future research objective can be study of different patterns which can transfer an idea to a patent. These knowledge transfer channels can be studies and compared while it can end to proposing a model to study effective idea to patent knowledge transfer patterns in citation networks. Also, in this research we were focused on directed citation graphs and studied cluster mappings based on this structure. As this study was the first regarding use of article-patent citations in one network, we suggest future studies on collaboration networks co-authorship and co-invertorship of this idea.

4.4 Chapter Summary

In this chapter, as the final chapter of the thesis, we concluded the research analyses results and discussed research limitations and future studies on citation networks of articles and patents in the Canadian nanotechnology

References

Aggarwal, C. C., and Zhai, C. (2012). A survey of text classification algorithms. In mining text data (pp. 163-222). Springer US.

Ahmad, K., and Al-Thubaity, A. (2003, July). Can text analysis tell us something about technology progress?, In Proceedings of the ACL-2003 workshop on Patent corpus processing-Volume 20 (pp. 46-55). Association for Computational Linguistics.

Al-Sultan, K. S., and Fedjki, C. A. (1997). A tabu search-based algorithm for the fuzzy clustering problem. Pattern Recognition, 30(12), (pp. 2023-2030).

Al-Sultan, K. S., and Khan, M. M. (1996). Computational experience on four algorithms for the hard clustering problem. Pattern recognition letters, 17(3), (pp. 295-308)

Al-Thubaity, A., and Ahmad, K. (2003, July). Knowledge maps as lexical signatures of journals papers and patent documents. In Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on (pp. 582-588).

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

Bassecoulard, E., Lelu, A., and Zitt, M. (2007). Mapping nanosciences by citation flows: A preliminary analysis. Scientometrics, 70(3), (pp. 859-880).

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

Beaudry, C., Schiffauerova A., (2011): Impacts of collaboration and network indicators on patent quality: The case of Canadian nanotechnology innovation. European Management Journal 29.5 (pp. 362-376).

Beaudry, C., and Schiffauerova, A. (2011). Is Canadian intellectual property leaving Canada? A study of nanotechnology patenting. The Journal of Technology Transfer, 36(6), (pp. 665-679).

Berkhin, P., (2006), Survey of Clustering Data Mining Techniques, Springer, pp. 25-71.

Bonaccorsi, Andrea, and Grid Thoma. (2007), Institutional complementarity and inventive performance in nano science and technology. Research Policy (pp. 813-831).

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

Brooks. H. (1994). The relationships between science and technology, Research Policy, Vol. 23, (pp. 477-486).

Canadian government reports (2012), Nanotechnologies international policy collaboration, See website: https://www.ic.gc.ca/eic/site/aimb-dgami.nsf/eng/h_03459.html

Carey C., Kaiser I. and Moore Valerie C., (2013). Nanotechnology patent literature review: Graphitic carbon-based nanotechnology and energy applications are on the rise. McDermott Will and Emery.

Charu C., Aggarwal. ChengXiang Z. (2012). A survey of text clustering algorithms. Springer. (pp. 77-128).

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

Chen, H. (2008). Mapping nanotechnology innovations and knowledge: global and longitudinal patent and literature analysis (Vol. 20). Springer Science and Business Media.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. Journal of the American Society for information Science and Technology, 57(3), (pp. 359-377).

Chen, C., and Hicks, D. (2004). Tracing knowledge diffusion. Scientometrics, 59(2), (pp. 199-211).

Compton V.J. (2004) Technological Knowledge: A developing framework for technology education in New Zealand. Briefing Paper prepared for the New Zealand Ministry of Education Curriculum Project.

Dang, Y., Zhang, Y., Fan, L., Chen, H., and Roco, M. C. (2010). Trends in worldwide nanotechnology patent applications: 1991 to 2008. Journal of Nanoparticle Research, 12(3), 687–706. http://doi.org/10.1007/s11051-009-9831-7

De Nooy, W., Mrvar, A., and Batagelj, V. (2011). Exploratory social network analysis with Pajek (Vol. 34). Cambridge Univ Pr. Blondel D. B., et. al. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment.

De Nooy, W., Mrvar, A., Batagelj, V. (2005), Exploratory Social Network Analysis with Pajek. Cambridge University Press.

De Solla Price, Derek J. (1965). "Is technology historically independent of science? A study in statistical historiography." Technology and Culture: (pp. 553-568).

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

DOE National Laboratories, (2007), Productive nanotechnology road-map, Battelle Memorial Institute and Foresight Nanotech Institute, Retrieved from: http://e-drexler.com/productivenanosystems/docs/Nanotech_Roadmap_2007_main.pdf

Dubin, D. (2004). The most influential paper Gerard Salton never wrote.

Dubes, R. C., and Jain, A. K. (1980). Clustering methodologies in exploratory data analysis. Advances in computers.

Easley D., and Kleinberg J., (2010), Networks, Crowds, and Markets: Reasoning about a Highly Connected World, Complete preprint on-line at http://www.cs.cornell.edu/home/kleinber/networks-book/

Egghe, L., and Rousseau, R. (1990). Introduction to informetrics: Quantitative methods in library, documentation and information science.

Erjia Yan EY. (2010), Applying centrality measures to impact analysis: A coauthorship network analysis. John Wiley and Sons.

Eslami, H., Ebadi A., and Schiffauerova, A. (2013), Effect of collaboration network structure on knowledge creation and technological performance: The case of biotechnology in Canada, Scientometrics 97.1: (pp. 99-119).

Freecite, (2009), developed by Brown University, implemented by Public Display and funded by the Mellon Foundation. Retrieved from: http://freecite.library.brown.edu/welcome

Fritz, M.; M. Radmacher, N. Petersen, H. E. Gaub (May 1994). Visualization and identification of intracellular structures by force modulation microscopy and drug induced degradation. The 1993 international conference on scanning tunneling microscopy. The 1993 international conference on scanning tunneling microscopy 12. Beijing, China: AVS. (pp. 1526–1529).

Glosary of patent law terms, (2015), Wikipedia: The Free Encyclopedia. Wikimedia Foundation, Inc., Web. 18 Aug. 2015. Retrieved from: https://en.wikipedia.org/wiki/Glossary_of_patent_law_terms#Non-patent_literature

Gridlogics Technologies, (2014), An analysis of patenting activity around 3D-Printing, Gridlogics Technologies Pvt Ltd, Website link: http://www.patentinsightpro.com/

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

Horikoshi, S., and Serpone, N. (Eds.). (2013). Microwaves in nanoparticle synthesis: fundamentals and applications. John Wiley and Sons.

Huang, Z., Chen, H., Yip, A., Ng, G., Guo, F., Chen, Z. K., and Roco, M. C. (2003). Longitudinal patent analysis for nanoscale science and engineering: Country, institution and technology field. Journal of Nanoparticle Research, 5(3-4), (pp. 333-363).

Haun, J. B., and Hammer, D. A. (2008). Quantifying nanoparticle adhesion mediated by specific molecular interactions. Langmuir, 24(16), (pp. 8821-8832).

Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P. J., Bunke, H., Goldgof, D. B., and Fisher, R. B. (1996). An experimental comparison of range image segmentation algorithms. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 18(7), (pp. 673-689).

Huang, C., Notten, A., and Rasters, N. (2011). Nanoscience and technology publications and patents: a review of social science studies and search strategies. The Journal of Technology Transfer, 36(2), (pp. 145-172).

Huang, Z., Chen, H., Chen, Z. K., and Roco, M. C. (2004). International nanotechnology development in 2003: Country, institution, and technology field analysis based on USPTO patent database. Journal of nanoparticle Research,6(4), (pp. 325-354).

Hullmann, A., and Meyer, M. (2003). Publications and patents in nanotechnology, Scientometrics, 58(3), (pp. 507-527).

Igami, M., and Saka, A. (2007). Capturing the evolving nature of science, the development of new scientific indicators and the mapping of science.

Jaffe, A. B., and Trajtenberg, M. (2002). Patents, citations, and innovations: A window on the knowledge economy. MIT press.

Jaffe, A. B., Trajtenberg, M., and Fogarty, M. S. (2000). The meaning of patent citations: Report on the NBER/Case-Western Reserve survey of patentees (No. w7631). National Bureau of Economic Research.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. ACM computing surveys (CSUR), 31(3), (pp. 264-323).

Jain, A. K., & Flynn, P. J. (1996). Image segmentation using clustering (pp. 65-83). IEEE Press, Piscataway, NJ.

Jordan, C. C., Kaiser, I. N., and Moore, V. C. (2012). Nanotechnology patent survey: who will be the leaders in the fifth technology revolution, Nano-tech. L. and Bus.

Jordan, C. C., Kaiser, I., and Moore, V. C. (2014). 2013 Nanotechnology Patent Literature Review: Graphitic Carbon-Based Nanotechnology and Energy Applications Are on the Rise. Nanotech. L. and Bus..

Kain A.K., Murty M.N., Flynn P.J. (1999). Data clustering: A review. ACM Computing Surveys. Volume 31 Issue 3, ages (pp. 264-323).

Kantardzic M. (2003), Data mining: Concepts, models, methods, and algorithms. John Wiley and Sons, Inc. New York, NY, USA

Karki, M. M. S. (1997). Patent citation analysis: A policy analysis tool. World Patent Information, 19(4), (pp. 269-272).

Karvonen, Matti, and Kässi. T., (2013), "Patent citations as a tool for analysing the early stages of convergence." Technological Forecasting and Social Change80.6: (pp. 1094-1107).

Kas, M. (2011). Structures and Statistics of Citation Networks. CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF ELECTRICAL AND COMPUTER ENGINEERING.

Kim, K.S.; Cota-Sanchez, German; Kingston, Chris; Imris, M.; Simard, Benoît; Soucy, Gervais (2007). "Large-scale production of single-wall carbon nanotubes by induction thermal plasma". Journal of Physics D: Applied Physics 40 (8): (pp. 2375–2387).

Kriebel, A., (2012), Stacked area chart vs. line chart – the Great Debate., http://vizwiz.blogspot.com/2012/10/stacked-area-chart-vs-line-chart-great.html

Kolleck N., (2013). Social network analysis in innovation research: Using a mixed, methods approach to analyze social innovations. Springer.

Kotsiantis S., Kanellopoulos D. and Pintelas P. (2006). Data preprocessing for supervised leaning. International Journal of Computer Science. 1(2):111.

Lawrence, S., and Giles, C. L. (1998). Searching the World Wide Web. Science, 280(5360), (pp. 98-100).

Lee, D. L., Chuang, Huei and Seamons, K., (1997), Document ranking and the vector-space model. IEEE Software. 14(2):67

Lee, R. C. T. (1981). Clustering analysis and its applications. In Advances in Information Systems Science (pp. 169-292). Springer US.

Leydesdorff, L., Wagner, C., Park, H. W., and Adams, J. (2013). International collaboration in science: The global map and the network. arXiv preprint arXiv:1301.0801.

Leydesdorff F., Loet, and Probst C. (2009). The delineation of an interdisciplinary specialty in terms of a journal set: The case of communication studies. Journal of the American Society for Information Science and Technology 60.8: (pp. 1709-1718).

Leydesdorff, L., Carley, S., and Rafols, I. (2013). Global maps of science based on the new Web-of-Science categories. Scientometrics, 94(2), (pp. 589-593).

Leydesdorff, L., and Rafols, I. (2009). A global map of science based on the ISI subject categories. Journal of the American Society for Information Science and Technology, 60(2), (pp. 348-362).

Leydesdorff, L. (2006). The knowledge-based economy: Modeled, measured, simulated. Universal-Publishers.

Li, X., Chen, H., Dang, Y., Lin, Y., Larson, C. A., and Roco, M. C. (2008). A longitudinal analysis of nanotechnology literature: 1976–2004. Journal of Nanoparticle Research, 10(1), (pp. 3-22).

Libaers, D., Meyer, M., and Geuna, A. (2006). The role of university spinout companies in an emerging technology: The case of nanotechnology. The Journal of Technology Transfer, 31(4), (pp. 443-450).

Lo, S. C. (2006). Linkage between public science and technology development of genetic engineering: Preliminary study on patents granted to Japan, Korea and Taiwan.

Lo, S. C. (2007). Patent analysis of genetic engineering research in Japan, Korea and Taiwan. Scientometrics, 70(1), (pp. 183-200).

Lo, Y. D., and Chiu, R. W. (2009). Next-generation sequencing of plasma/serum DNA: an emerging research and molecular diagnostic tool. Clinical chemistry, (pp. 607-608).

Lopez, J., Potter, G., (2001). After Postmodernism: An Introduction to Critical Realism, AandC Black, ISBN: 1847141064, 9781847141064

Marienville, S (1992), Entrepreneurship in the Business Curriculum, Journal of Education for Business, Vol. 68 No. 1, (pp. 27-31).McVeigh, M. E. (2004). Open access journals in the ISI citation databases: analysis of impact factors and citation patterns: a citation study from Thomson Scientific (p. 125). Thomson Scientific.

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

Meyer, M. (2007). What do we know about innovation in nanotechnology? Some propositions about an emerging field between hype and path-dependency. Scientometrics, 70(3), (pp. 779-810).

Meyer, M. (2006). Knowledge integrators or weak links? An exploratory comparison of patenting researchers with their non-inventing peers in nano-science and technology. Scientometrics, 68(3), (pp. 545-560).

Meyer, M. (2006). Are patenting scientists the better scholars?: An exploratory comparison of inventor-authors with their non-inventing peers in nano-science and technology. Research Policy, 35(10), (pp. 1646-1662).

Meyer, M., Debackere, K., and Glänzel, W. (2010). Can applied science be 'good science'? Exploring the relationship between patent citations and citation impact in nanoscience. Scientometrics, 85(2), (pp. 527-539).

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

Narin, F., Hamilton, K. S., and Olivastro, D. (1997). The increasing linkage between US technology and public science. Research Policy, 26(3), (pp. 317-330).

National Science Foundation, (2001) Societal implications of nanoscience and nanotechnology. Arlington, VA (Also published by Kluwer Academic Publishing, 2001). See NSF website: http://www.nsf.gov/crssprgm/nano/reports/ nsfnnireports.jsp

Nelson, R.R., (1992). National innovation systems: A retrospective on a study, Industrial and Corporate Change 1, (21): (pp. 347-374).

Nelson, R.R. and N. Rosenberg, (1993). Technical innovation and national systems, in R.R. Nelson (editor), National Innovation Systems: A Comparative Analysis (Oxford University Press, New York/Oxford), (pp. 1-21).

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

Noyons, E. C. M., Buter, R. K., Van Raan, A. F. J., Schmoch, U., Heinze, T., Hinze, S., and Rangnow, R. (2003). Mapping excellence in science and technology across Europe: life sciences. Nanoscience and nanotechnology.

Perez C., (2013). Technological revolutions and financial capital. Edward Elgar Pub.

O'Brien, N. J., and Cummins, E. J. (2011). A risk assessment framework for assessing metallic nanomaterials of environmental concern: aquatic exposure and behavior. Risk Analysis, 31(5), (pp. 706-726).

Owen-Smith, J., and Powell, W. W. (2008). Networks and institutions. The Sage handbook of organizational institutionalism, (pp. 596-623).

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

Pyle D. (1999). Data preparation for data mining. Morgan Kaufmann Publishers Inc.

Raghavan, V. V., Wong, S. K. M. A. (1986): A critical analysis of vector space model for information retrieval. Journal of the American Society for Information Science; 37(5).

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

Rocco, T. S., andand Plakhotnik, M. S. (2009). Literature reviews, conceptual frameworks, and theoretical frameworks: Terms, functions, and distinctions. Human Resource Development Review.

Ruda, H., (2008), The center for advanced nanotechnology, copyright 2001, 2008 center for advanced nanotechnology university of Toronto. Retrieved from: http://sites.utoronto.ca/ecan/index.html

Saka, A., andand Igami, M. (2007, July). Mapping modern science using co-citation analysis. In Information Visualization, 2007. IV'07. 11th International Conference (pp. 453-458). IEEE.

Salton G., Wong A. and Yang C. S. (1975): A vector space model for athematic indexing. Communications of the ACM; 18(11):613.

Small H., (1973), Co-citation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for Information Science: 24.

Sanderson, K. (2006). Sharpest cut from nanotube sword. Nature news.

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

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

Shiffrin, R. M., and Börner, K. (2004). Mapping knowledge domains. Proceedings of the National Academy of Sciences, 101(suppl 1), (pp. 5183-5185).

Tahmooresnejad, L., Beaudry, C., and Schiffauerova, A. (2015). The role of public funding in nanotechnology scientific production: Where Canada stands in comparison to the United States. Scientometrics, 102(1), (pp. 753-787).

Thomson Reuters, (2008), A guide to evaluating research performance with citation data, Accessible by: http://ip-science.thomsonreuters.com/m/pdfs/325133_thomson.pdf

Van Raan, A., and Tijssen, R. (1993). The neural net of neural network research: An exercise in bibliometric mapping. Scientometrics, 26(1), (pp. 169-192).

Verbeek, A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E., and Deleus, F. (2002). Linking science to technology: Using bibliographic references in patents to build linkage schemes. Scientometrics, 54(3), (pp. 399-420).

Vieira, E., and Gomes, J. (2009). A comparison of Scopus and Web of Science for a typical university. Scientometrics, 81(2), (pp. 587-600).

Wasserman S., Faust. (1994). Social network analysis: Methods and applications (structural analysis in the social sciences); 978-0521387071.

Wong, P. K., Ho, Y. P., and Chan, C. K. (2007). Internationalization and evolution of application areas of an emerging technology: The case of nanotechnology.Scientometrics, 70(3), (pp. 715-737).

Tijssen, R. J., andand Van Raan, A. F. (1994). Mapping Changes in Science and Technology Bibliometric Co-Occurrence Analysis of the R&D Literature.Evaluation Review, 18(1), (pp. 98-115).

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

Yan E., Ding Y., (2009). Applying centrality measures to impact analysis: A coauthorship network analysis, Journal of the American Society for Information Science and Technology

Yan, E., Ding, Y., and Zhu, Q. (2010). Mapping library and information science in China: A coauthorship network analysis. Scientometrics, 83(1), (pp. 115-131).

Yan, E., Ding, Y., and Sugimoto, C. R. (2011). P-Rank: An indicator measuring prestige in heterogeneous scholarly networks. Journal of the American Society for Information Science and Technology, 62(3), (pp. 467-477).

Yan, E., andand Ding, Y. (2011). Discovering author impact: A PageRank perspective. Information processing and management, 47(1), (pp. 125-134).

Yan, E., Ding, Y., andand Jacob, E. K. (2011). Overlaying communities and topics: an analysis on publication networks. Scientometrics, 90(2), (pp. 499-513).

Ynalvez, M. A., andand Shrum, W. M. (2011). Professional networks, scientific collaboration, and publication productivity in resource-constrained research institutions in a developing country. Research Policy, 40(2), (pp. 204-216).

Appendix-A: List of Technological and Scientific Clusters

Technological Clusters
Carbon nanotubes
Thin films
Electric field
Vapor decompositions
Atomic force microscopy
Synthesis and characterization
Aligned carbon
Electromagnetic field
Scanning tunneling microscopy
Optical properties
Preparation and characterization
Gold nanoparticles
Silicon nanowires
Quantum Computing
Drug delivery
Nanowire arrays
Block copolymer
Silver nanoparticles
Magnetic properties
Gene delivery
Scanning probe microscopy
Porous silicon
Atomic layer
Low temperature

Scientific Clusters
Carbon nanotubes
Composite material
Force microscope
Scanning probe microscopy
Atomic force
Nucleic acid
Particle size
Surface area
Metal oxide
Drug delivery
Magnetic field

Vapor decompositions
Delivery systems
Thin film
Electric field
Organic solvent
Semiconductor nanocrystal
Nanotube film
Optical devices
Medical devices
Refractive index
Scanning probe microscopy
Plurality of nanoparticles
Pharmaceutical agent
Block copolymer
Transition metal
Fuel cell
Contrast agent
Aqueous solution
Biocompatible polymer
Present disclosure
Metal ions
Optical fiber
Solid phase
Gas phase
Phase change
Magnetic resonance
Electron beam
Microcrystalline cellulose
Gas sensor
Chemical species
Carbon dioxide
Reaction chamber
Calcium phosphate
Electromagnetic field