6-16-2014

Social Network Analysis of Researchers' Communication and Collaborative Networks Using Self-reported Data

Oguz Cimenler
University of South Florida, oguzcimenler@gmail.com

Social Network Analysis of Researchers' Communication and Collaborative Networks Using Self-reported Data

by

Oguz Cimenler

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Industrial and Management Systems Engineering
College of Engineering
University of South Florida

Major Professor: Kingsley A. Reeves, Jr., Ph.D.
José Zayas-Castro, Ph.D.
Alex Savachkin, Ph.D.
Adriana Iamnitchi, Ph.D.
John Skvoretz, Ph.D.

Date of Approval:
June 16, 2014

Keywords: Informetrics, individual innovativeness, exponential random graph models (ERGMs), Poisson regression analysis, structural equation modeling

DEDICATION

I dedicate this dissertation to my adorable wife, Ummuhan Cimenler, who was always supportive to me during my study, to my beloved parents, Cumali and Emine Cimenler, particularly to my father, Cumali Cimenler, who is currently struggling with lung cancer, and to my dear brother, Omer Cimenler.

ACKNOWLEDGMENTS

I cannot express the degree of my gratitude to my adviser, Dr. Kingsley A. Reeves, who has encouraged and supported me at the times that I most needed. To prosper in my study, he always challenged me in a positive and constructive way, which was neither obstructive nor demanding.

I am very fortunate to have such an important scholar in the field of Social Network Analysis as Dr. John Skvoretz, on my dissertation committee. He was always helpful and guided me with his key and fruitful comments about my research. I would like to thank Dr. Alex Savachkin, Dr. Jose Zayas-Castro, and Dr. Adriana Iamnitchi for serving on my research committee and giving me valuable comments and feedbacks during the development stages of my research. Especially, I would like to make special thank you to Dr. Alex Savachkin for his motivational support during course of my study.

I also would like thank to all of tenured and tenured-track faculty members in the University of South Florida's College of Engineering. They have dedicated their time, energy and effort by filling out to my questionnaire which was one of the core tools I used in my study. Particularly, I would like to thank to Dr. Ali Yalcin and Dr. Gokhan Mumcu for their feedbacks and comments during the pilot-study stage of my questionnaire. Finally, I would like to thank to Dr. John Kuhn for being the chair of my dissertation defense meeting and Ms. Catherine Burton and Dr. Anna Dixon for her patience and corrections on my dissertation draft.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This research seeks an answer to the following question: what is the relationship between the structure of researchers' communication network and the structure of their collaborative output networks (e.g. co-authored publications, joint grant proposals, and joint patent applications), and the impact of these structures on their citation performance and the volume of collaborative research outputs? Three complementary studies are performed to answer this main question as discussed below.

1. Study I: A frequently used output to measure scientific (or research) collaboration is co-authorship in scholarly publications. Less frequently used are joint grant proposals and patents. Many scholars believe that co-authorship as the sole measure of research collaboration is insufficient because collaboration between researchers might not result in co-authorship. Collaborations involve informal communication (i.e., conversational exchange) between researchers. Using self-reports from 100 tenured/tenure-track faculty in the College of Engineering at the University of South Florida, researchers' networks are constructed from their communication relations and collaborations in three areas: joint publications, joint grant proposals, and joint patents. The data collection: 1) provides a rich data set of both researchers' in-progress and completed collaborative outputs, 2) yields a rating from the researchers on the importance of a tie to them 3) obtains multiple types of ties between researchers allowing for the comparison of their multiple networks. Exponential Random Graph Model (ERGM) results show that the more communication researchers have the more

likely they produce collaborative outputs. Furthermore, the impact of four demographic attributes: gender, race, department affiliation, and spatial proximity on collaborative output relations is tested. The results indicate that grant proposals are submitted with mixed gender teams in the college of engineering. Besides, the same race researchers are more likely to publish together. The demographics do not have an additional leverage on joint patents.

2. Study II: Previous research shows that researchers' social network metrics obtained from a collaborative output network (e.g., joint publications or co-authorship network) impact their performance determined by g-index. This study uses a richer dataset to show that a scholar's performance should be considered with respect to position in multiple networks. Previous research using only the network of researchers' joint publications shows that a researcher's distinct connections to other researchers (i.e., degree centrality), a researcher's number of repeated collaborative outputs (i.e., average tie strength), and a researchers' redundant connections to a group of researchers who are themselves well-connected (i.e., efficiency coefficient) has a positive impact on the researchers' performance, while a researcher's tendency to connect with other researchers who are themselves well-connected (i.e., eigenvector centrality) had a negative impact on the researchers' performance. The findings of this study are similar except that eigenvector centrality has a positive impact on the performance of scholars. Moreover, the results demonstrate that a researcher's tendency towards dense local neighborhoods (as measured by the local clustering coefficient) and the researchers' demographic attributes such as gender should also be considered when investigating the impact of the social network metrics on the performance of researchers.
3. Study III: This study investigates to what extent researchers' interactions in the early stage of their collaborative network activities impact the number of collaborative outputs produced

(e.g., joint publications, joint grant proposals, and joint patents). Path models using the Partial Least Squares (PLS) method are run to test the extent to which researchers' individual innovativeness, as determined by the specific indicators obtained from their interactions in the early stage of their collaborative network activities, impacts the number of collaborative outputs they produced taking into account the tie strength of a researcher to other conversational partners (TS). Within a college of engineering, it is found that researchers' individual innovativeness positively impacts the volume of their collaborative outputs. It is observed that TS positively impacts researchers' individual innovativeness, whereas TS negatively impacts researchers' volume of collaborative outputs. Furthermore, TS negatively impacts the relationship between researchers' individual innovativeness and the volume of their collaborative outputs, which is consistent with 'Strength of Weak Ties' Theory. The results of this study contribute to the literature regarding the transformation of tacit knowledge into explicit knowledge in a university context.

CHAPTER 1: INTRODUCTION

1.1. Statement of the Research Problem

A Science and Technology (S&T) system comprises a wide range of activities such as fundamental science or scholarly activity, and applied research and developmental activities mainly concentrating on creating new products and processes [1]. S&T system has become a driving force over the last 20 years for major economic growth and development, and it is, therefore, an inseparable part of several national and regional innovation systems [1, 2]. Innovation is one of the principal drivers of today's competitiveness [3]. As mentioned in a strategy report prepared by the White House, "America's economic growth and competitiveness depends on its people's capacity to innovate" [4]. However, competitive disadvantages can be turned into advantages through collaboration [5]. Therefore, it is important to establish a balance between conflicting goals such as competition and collaboration [3]. Furthermore, innovation has three dimensions that need to be taken into account: human dimension such as talent for knowledge creation, financial dimension such as governmental funding, and infrastructural dimension such as policy generation for building networks between different entities [3].

One of the important attributes contributing to the S&T system performance is scientific collaboration [1, 6]. Sonnenwald (2007) defined scientific collaboration as the interaction within a social context among two or more scientists in order to facilitate the completion of tasks with regard to a commonly shared goal. Thus, participants in the collaboration event integrate valuable knowledge from their respective domains to create new knowledge. Scientific collaboration provides several salient advantages, for example; 1) access to expertise for complex problems, new resources and, funding [6-13], 2) increase in the participants' visibility and recognition [8, 10], 3) rapid solutions for more encompassing problems by creating a synergetic effect among participants [10, 14], 4) decrease in the risks and possible errors made, thereby increasing accuracy of research and quality of results due to multiple viewpoints [10, 11], 5) growth in advancement of scientific disciplines and cross-fertilization across scientific disciplines [10, 15], 6) development of the scientific knowledge and technical human capital, e.g., participants' formal education and training, and their social relations and network ties with other scientists [16], 7) increase in the scientific productivity of individuals and their career growth [8, 16-18], and 8) help in extending the scope of a research project and fostering innovation since additional expertise is needed [7]. One of the important factors leading to advantages of scientific collaboration is the social dimension of scientific work such as informal conversational exchanges between colleagues [8, 16], co-authorship relations [8, 19], jointly submitted grant proposals [8, 20], co-patent applications [21-24]. To be able to develop a greater collaboration among individual researchers, which leads to these salient advantages, and to formulate policies that aim at improving the relationships between researchers, it is necessary to investigate the relationship between the structure of their communication network and the structure of researchers' collaborative output networks (e.g. co-authored or joint publications, joint grant proposals, and joint patents), and the impact of these structures on their citation performance and the volume of collaborative research outputs. In addition, analyzing these networks and their relationship with researchers' performance and the volume of collaborative research outputs contributes to our understanding regarding the infrastructural dimension of innovation.

Co-authorship in scholarly publications is the most tangible and well-documented forms of scientific collaboration, and it is also a good indicator of the S&T system performance. Therefore, it is used widely in scientific collaboration studies [1, 8, 14, 19]. For example, using social network analysis (SNA), Newman [25-27] and Barabási et al. (2002) analyzed the structural properties of scientific collaboration patterns in large scale by depicting the network of researchers when two authors were considered linked if their names appeared in the same scientific journal. They found that co-authorship networks were small world networks in which most nodes (i.e., authors) could be reached from other nodes by a small number of steps. With a similar approach used in co-authorship network studies, some studies also analyzed the structure of co-inventor maps in the case that two patent applicants (i.e. co-authors) were linked if there was a patent application together by these two applicants; thus, a network of co-invention was constructed. However, analyzing co-inventor maps was not used as widely as analyzing co-authorship maps [22]. In addition, for the networks constructed from researchers' jointly submitted grant proposals, there was not to my knowledge any study in the literature analyzing the properties of these networks, their relations with other concepts, and concomitant implications.

Many scholars argue that co-authorship alone is insufficient as a measure of research collaboration. For example, Katz and Martin (1997) pointed out that many cases of collaboration did not result in co-authored publications; for example when researchers worked closely together but decided to publish their results separately due to the fact that they came from different fields and desired to produce single-author papers in their own discipline. Their study concluded that measuring co-authorship was a partial indicator of research collaboration. Melin and Persson (1996) also asserted that co-authorship was only a rough indicator of collaboration, even though significant scientific collaboration leads to coauthored publications in most cases. The qualitative study of Laudel (2002) determined different types of collaborations that were classified according to the content of contribution made by collaborators. Then, a collaborator was rewarded with a co-authorship depending on the level of his/her contribution. The assumption that co-authorship and research collaboration are synonymous was criticized by several other scholars for the following reasons: listing co-authors for purely social reasons [8, 16, 30], listing co-authors simply by the virtue of providing material or performing a routine task [8, 16, 31], making the colleagues 'honorary co-authors' [8, 16, 32], and listing co-authors who did not even communicate with each other during research collaboration (e.g., many publications in physics and astrophysics include hundreds of authors) [33].

Fox (1983) stated that communication and exchange of research findings and results were the most fundamental social process of science, and the principal means of this communication was the publication process. Communication between researchers not only stimulates them to think regarding the unsolved problems in their fields and possible research projects, thereby developing new ideas and solutions, but it also transmits ‘know-how’ or the procedural knowledge to efficiently solve the problems to other researchers [29]. Collaborations mostly begin informally and arise from informal communication between researchers, i.e., through close personal contacts and professional networks [8, 16, 30, 34-36]. Kraut and Edigo (1988) found that researchers in a close physical proximity tended to collaborate more due to the changes in three properties of informal communication: increasing the frequency of communication, increasing the quality of communication, and reducing the cost of communication. Olson and Olson (2000) also reported that face-to-face communication facilitates the flow of situated cognitive and social activities due to some of its key characteristics such as rapid feedback and multiple channels (e.g, voice, facial expression, gesture, body posture). However, the use of information and communication technologies (ICT) such as audio and video conferences, mobile phones, e-mail, social networking sites especially designed to support collaborative environment, and the World Wide Web facilitate informal communication between researchers and help them collaborate with other distant researchers in a timely manner [7, 39, 40]. Using both types of communication, face-to-face and ICT, have their own advantages and disadvantages [38]. In sum, communication is an important source and influential factor for scientific collaboration [6, 8, 11, 41] and a fundamental component to sustain collaboration [7].

Many scholars make a clear distinction between researchers' communication and collaboration. For example, Melin and Persson (1996) reported that "collaboration was an intense form of interaction that allowed for effective communication". Melin (2000) discussed that collaboration could be measured in a number of ways such as exchange of phone calls and e-mails, but a more concrete form to measure the collaboration was through co-authorship information. Laudel (2002) accepted publications as a way of formal communication, and found out that a considerable proportion of collaborations were not rewarded as a co-authorship. Borgman and Furner (2002) discussed that collaboration was one of the communication behaviors exhibited by authors in their various capacities. Similarly, from a network viewpoint, Newman, 2001b reported that there was an assumption that most people who wrote a paper together might not be genuinely acquainted with one another. Consequently, even though there is a clear distinction between researchers' communication and collaboration, considering the researchers' communication and collaborative output networks separate from each other is not fully addressed in the literature. Taking the assumption reported by Newman (2001b), one notable study made by Pepe (2011) compared the structure of researchers' communication network with the structure of their collaborative output network (e.g., co-authorship network) by utilizing techniques used in SNA. The study found out the extent to which the structure of researchers' communication network overlaps the structure of their collaborative output network. That is, the more these network structures overlap the more likely collaborative output relations between researchers can be seen as a surrogate or proxy for communication relations between researchers.

Analyzing scientific collaboration through co-authorship indicator is performed at micro (individual) level, meso (institutional) level, and macro (national, international, and multinational) [19, 41]. The knowledge at meso and macro level did not yet adequately reflect the trends in cooperation between researchers; therefore, there should be more efforts to investigate collaboration at micro level which is the lowest level of aggregation [41-43]. Hence, SNA is the promising method to investigate the trends in cooperation and reveal the structure of collaboration between individuals [42, 43]. In addition, collaboration is related to many types of shared attributes [16, 30]; therefore, these four networks should be analyzed by taking some demographic attributes of individuals such as gender, race, departmental affiliation, and spatial proximity into consideration. In the light of the above discussion, this study mainly addresses four issues in the literature:

1. The case that co-authorship is seen as the partial or rough indicator of scientific collaboration.
2. The degree to which researchers' collaboration network can be regarded as a proxy for their communication network.
3. The extent to which researchers' communication network impacts their collaboration networks.

4. The comparative analysis of the researchers' multiple networks which are constructed by the researchers' communication ties (i.e. conversational exchange ties) and their collaborative output ties (e.g., co-authored or joint publications, joint grant proposals, and joint patents) with other researchers.

The first issue can be addressed by extending an existing data collection method, which is already used for collecting the number of researchers' collaborative output with other researchers via self-report, into the social network context. Even though the second issue has already been addressed by Pepe (2011), I extend this into researchers' multiple networks which are constructed from a dataset obtained through via researchers' self-report. As previously discussed, communication among researchers initiates their collaborative activities. However, to what extent that the structural aspect of researchers' communication relations impacts the structural aspect of researchers' collaboration relations is not fully addressed in the literature. Thus, by addressing the third issue, this study's findings also have the ability to measure the extent to which collaboration among researchers is nurtured by means of their conversational exchange in the network context. The fourth issue is because there is a major limitation in gathering data with regard to a researcher's communication as well as collaborative output information with other researchers (see next section for further discussion). To overcome this major limitation, the relational data for researchers' multiple networks (e.g., researchers' communication and collaborative output networks) can be simultaneously collected at either the individual college level or at the university as a whole.

1.2. Proposed Solution

Considering the discussion in the literature that relying solely on co-authorship relations is not a sufficient indicator of scientific collaboration, Bozeman and Corley (2004) and Lee and

Bozeman (2005) employed participants' self-report of collaboration information, which permitted the participants to indicate which relationships are worthy of being considered as collaborations. Using a questionnaire, they asked participants to make a self-report of the number of people with whom they had engaged in research collaborations within the past 12 months. Referring to the past literature, they discussed that the self-reported way of collecting the collaboration data avoided some of the problems seen in the publication-based measure of collaboration, for instance, listing the authors purely for social reasons [8, 30], listing the authors for simply providing material or performing a routine task [8, 31], and making colleagues honorary co-authors [8, 32]. Even though Lee and Bozeman (2005) and Vasileiadou (2009) highlighted the disadvantages of the self-reported way of collecting data such as accuracy of the collected data, there are many recent studies using the method of collecting collaboration information via self-report [45-48].

Their method can be extended to collecting researchers' communication and collaboration information in a social network context by employing a questionnaire where researchers identify their contacts and provide the amount of communication and collaboration with those contacts via self-reports. For example, while collecting the collaboration information, a participant can be asked to report the names of the researchers with whom he/she has engaged in both communication and research collaborations together with the frequency of that communication and the number of collaborative outputs (both in-progress and completed) with those reported names via a name generator. By reporting both of their in-progress and completed collaborative output ties (e.g., co-authorship ties), they can decide on which ties are important to them and whether or not reported contact is actually involved in research. This helps overcome the challenge that many collaborations do not result in tangible outcome such as co-authorship by capturing in-progress collaborative output ties as well as other challenges, such as co-authors who are listed for only social reasons and co-authors that are not even communicated. It will be more successful if this method can be executed within the college of a university or even within a university because close proximity of the researchers will facilitate data collection in a way that the relational data for mapping the researchers' multiple networks (e.g., network of communications, network of joint publications, grant proposals, and patents) can be simultaneously collected at either the individual college level or at the university as a whole. Moreover, the name generator can contain prepopulated names of the researchers within the college of a university in order to help the participant for ease of remembering the names.

In addition to abovementioned advantages, administering a self-reported questionnaire can overcome the major limitation in gathering data with regard to a researcher's communication as well as collaborative output information with other researchers. The limitation is mainly due to these challenges: the unavailability of data for multiple networks, the inability to access the multiple data repositories, and the difficulty of scanning multiple databases. For example, for the same researchers, data might be available and easily accessible in order to construct the network of co-authorships or joint publications, but either unavailable or difficult to access in order to construct the network of communications, joint grant proposals, and patents. Moreover, scanning the different databases to collect the same researchers' both communication and collaborative output information might also be tedious job.

In sum, the self-reported way of collecting data provides the following benefits:

- Researchers can be asked to report both communication and collaborative output (in-progress and completed) information with other researchers.

- Researchers assess which ties are important to them according to their own perceptions and whether or not reported contact is actually involved in research.
- Relational data for multiple networks (e.g., researchers' network of communications and collaborative outputs) is simultaneously collected.

1.3. Statement of Research Objectives

This study focuses on the population of research faculty within the University of South Florida's College of Engineering. Data was collected by employing a questionnaire by which researchers report their contacts, the number of collaborative outputs, and the frequency of communication with them in a self-reported manner. The relational data obtained through the questionnaire was put into the form of a two-way matrix where rows and columns referred to researchers making up the pairs [49]. Furthermore, each cell in the matrix indicated the collaborative output or communication ties between the researchers. Thus, four 100x100 matrixes were constructed from the relational data provided by the researchers: a matrix of communication relations and a matrix of joint publications (or co-authorship), grant proposals, and patents. Using the relational data, the first objective of this study is:

1. To investigate how similar are researchers' communication network and collaborative output networks (i.e. joint publications, grant proposals, and patents) and what is the impact of the communication network structure on the structure of collaborative output networks in the presence of demographic attributes.

To be able to accomplish the first objective, several sub-objectives which require visualization of the networks and further statistical analyses are fulfilled. These sub-objectives are follows:

a. To examine the statistical and descriptive properties of these four networks (i.e. network of communications, co-authored publications, joint grant proposals, and joint patent applications).
b. To investigate the correlation between the ties that are present in one network and the ties that are present in others.
c. To explore how the presence of a tie in the communication network from one researcher to another would increase the likelihood of the presence of a tie in each collaborative output network.
d. To investigate whether or not the structural location of an individual or a similar group of individuals is advantageous across the four networks.
e. To investigate if the sharing of some attribute by two researchers facilitates tie formation between them across the four networks (i.e., homophily hypothesis).
f. To investigate whether or not researchers who have a similar spatial proximity tend to produce collaborative outputs together.

The quality of research outputs is as important as the quantity of the research outputs. Hirsch (2005) proposed an index called the h-index in order to attempt to measure both the number of publications a researcher produced (i.e., quantity) and their impact on other publications (i.e., quality). Using the researchers' publications data in the information schools of five universities, Abbasi et al. (2011) investigated the impact of social network metrics (including different centrality metrics, average tie strength, and efficiency coefficient proposed by Burt (1992)) obtained from a researchers' co-authorship network on the their g-index (another form of h-index). Their study can be extended by considering the network metrics obtained from researchers' multiple networks. Using the data gathered by the questionnaire, the second objective is the following:

2. To test the impact of social network metrics extracted from both researchers' communication and collaborative output networks (e.g., degree, closeness, betweenness, and eigenvector centralities, average tie strength, and efficiency coefficient proposed by Burt (1992), local clustering coefficient) on the researchers' citation-based performance index (h-index).

Bjork and Magnusson (2009) asserted that "innovation can be seen as ideas that have been developed and implemented". Working as a group and attending to the ideas of the others could both spark a good idea and lead to a novel combination of ideas. Then, collaboration is necessary for creativity, innovation, and problem solving [54, 55]. From the network perspective, Lovejoy and Sinha (2010) found that individual innovativeness during the ideation phase was accelerated by two properties: 1) an individual's participation in a 'maximal complete sub-graph' or clique (called just 'complete graphs' in their study), which maximizes the number of parallel conversations, and 2) the knowledge gain of individuals via their conversational churn which means that an individual constantly changes his/her conversational partners through a large set of conversational partners. In addition to these two properties, perceived self-innovativeness should also be considered as an accelerator of the individual innovativeness [57-62]. In the literature, investigating the relationship between researchers' individual innovativeness during ideation phase and their collaborative outputs is not addressed. This is because the studies in the literature mostly focus on the final outputs such as publications and citations due to the major limitation of collecting information with regard to researchers' interaction in early stages of their collaborative activities. It is also important to consider the tie strength of a researcher to other conversational partners while investigating the relationship between researchers' individual innovativeness and their collaborative outputs because knowledge creation is an important step which supports idea generation [63] and the strength of an interpersonal connection impacts how easily the created knowledge can be transferred to other individuals [64-67].

To investigate the relationship between researchers' individual innovativeness and their collaborative outputs taking into account the tie strength of a researcher to other conversational partners, the path model with three latent variables-LVs, shown in Figure 1.1, is proposed to test the four hypotheses. A path model consists of different latent variables-LVs (also called unobservable variables, constructs, and factors) and their related indicators or observable variables [68]. The LV, researchers' individual innovativeness, has three indicators: researchers' rate of participation in 'complete graph(s)' [56], researchers' knowledge gain via their conversational churn [56, 69], and the perceived self-innovativeness score of researchers [57, 59]. The LV, collaborative outputs, has three indicators: the number of researchers' collaborative outputs such as joint publications, grant proposals, and patents. The LV, tie strength of an individual to others, has three indicators: frequency of interaction called 'frequency', 'closeness', and 'intimacy' (or mutual confiding) with conversational partners [70, 71]. Therefore, third objective is the following:

3. To test the impact of researchers' individual innovativeness (as determined by the specific indicators obtained from their communication network) on the volume of their collaborative outputs taking into account the tie strength of a researcher to other conversational partners.

To be able to accomplish this objective, below sub-objectives need to be fulfilled:

   a. To test the impact of researchers' individual innovativeness on the volume of their collaborative outputs.

b. To test the impact of tie strength of an individual to others on both researchers' individual innovativeness and the volume of their collaborative outputs.

c. To test the moderating effect of tie strength of an individual to others on the impact of researchers' individual innovativeness on the volume of their collaborative outputs.

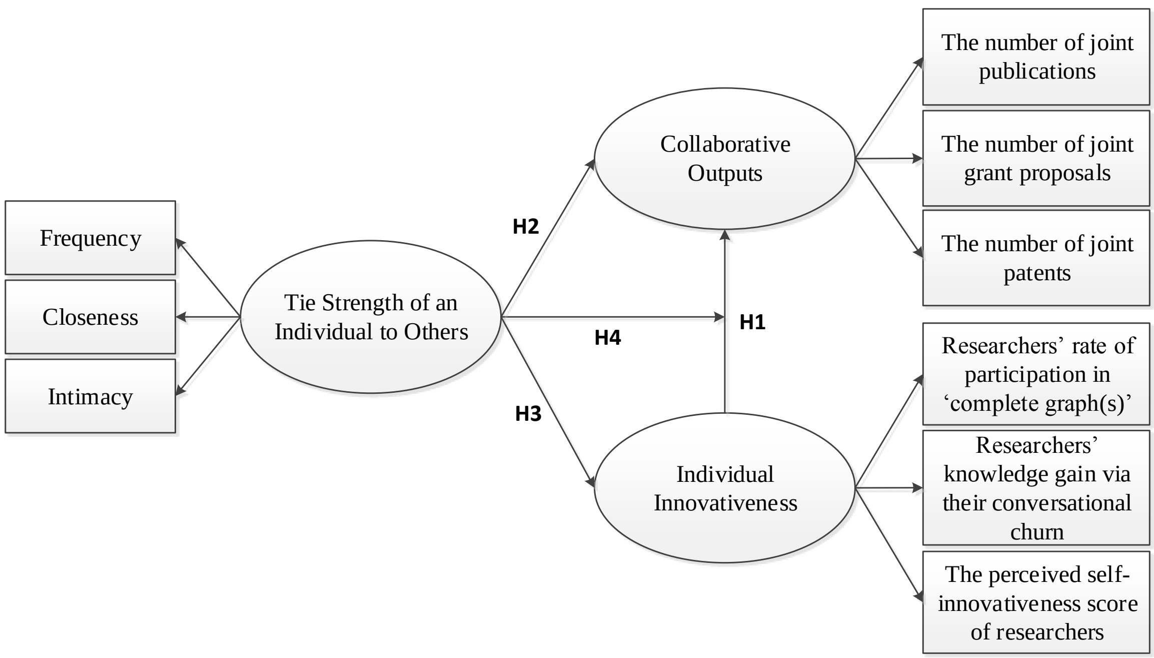

Figure 1.1. Path Model for the Third Research Objective

CHAPTER 2: AN EVALUATION OF COLLABORATIVE RESEARCH IN A COLLEGE OF ENGINEERING: A SOCIAL NETWORK APPROACH

2.1. Introduction

The most frequently used output to measure research collaboration is co-authorship in scholarly publications. However, many scholars discussed that co-authorship is an insufficient singular measure of research collaboration. The reason for this is threefold: 1) not all collaborations resulted in co-authored publications, 2) authors might be listed in publications for purely social reasons such as ‘honorary coauthors’, and 3) authors appearing in the same publication sometimes do not communicate with each other [8, 14, 29, 33]. With a similar approach used in co-authorship network studies, some studies also analyzed the structure of co-inventor maps in the case that two patent applicants (i.e. co-authors) were linked if there was a patent application together by these two applicants; thus, a network of co-invention was constructed. However, analyzing co-inventor maps is not used as widely as analyzing co-authorship maps [22]. In addition, for the networks constructed from researchers’ jointly submitted grant proposals, there was not to my knowledge any study in the literature analyzing the properties of these networks, their relations with other concepts, and related implications.

Collaborations mostly arise from informal communication between researchers [8, 16, 30, 34-36]. Therefore, many scholars make a clear distinction between researchers’ communication and collaboration [9, 14, 39]. Despite these aforementioned facts, the relationship between researchers’ communication network and collaborative output networks (e.g., network of co-authored or joint publications, grant proposals, and patents) in which a tie between any two authors indicates collaboration on the making of a collaborative output, and the impact of the former on the latter in the presence of some demographic attributes (e.g., gender, race, department affiliation, and spatial proximity) are not fully addressed in the literature. Then, to be able to develop a greater collaboration among individual researchers, and to formulate policies that aim at improving the relationships between researchers, it is necessary to investigate this literature gap. Considering abovementioned multiple networks constructed by self-reports from 100 tenured and tenured-track faculty in the College of Engineering at the University of South Florida, this chapter seeks an answer for the following question: *How similar or dissimilar are researchers' communication network and collaborative output networks (i.e. joint publications, grant proposals, and patents) and what is the impact of the communication network structure on the structure of collaborative output networks in the presence of demographic attributes?*

2.2. Literature Review and Hypotheses

2.2.1. The Field of Informetrics

The field of informetrics or only 'informetrics' studies the quantitative aspects of information in any form, not only just records or bibliographies but also informal or spoken communication, and in any social group, not just scientists, and already started in the first half of the twentieth century [72, 73]. Informetrics now is a broader and general term which is comprised of studies related to information science such as bibliometrics, scientometrics, webometrics, and cybermetric [73-75]. Bibliometrics studies the quantitative aspects of the production, dissemination, and use of recorded information such as scientific papers, articles, and books [72-74]. Scientometrics studies the quantitative aspects of science as a discipline or economic activity and mostly deals with science policy, citation analysis, and research evaluation

[72, 73]. Webometrics studies the quantitative aspects of the construction and use of information resources, structures and technologies on the Web in four main areas: web page content analysis, web link structure analysis, web usage analysis (including log files of users' searching and browsing behavior), and web technology analysis (including search engine performance) [76]. Since many scholar activities are becoming web-based, webometrics are covered by bibliometrics and scientometrics to some extent [76]. Cybermetrics studies the quantitative aspects of the construction and use of information resources, structures, and technologies on the whole Internet, e.g., including statistical studies of discussion groups, mailing lists, and other computer mediated communication on the web [76]. Scientific collaboration measured by co-authorship relations is a classical subfield of informetrics and mostly connected to bibliometric and scientometrics studies [73, 75]. Therefore, in the field of informetrics, there are many studies devoted into collaboration patterns and relationships between researchers by constructing collaboration networks at author level [75].

2.2.2. Scientific Collaboration

A science and technology (S&T) system comprises a wide range of activities such as fundamental science or scholarly activity, and applied research and developmental activities mainly concentrating on creating new products and processes [1]. It has become a driving force over the last 20 years for major economic growth and development and it is, therefore, an inseparable part of several national and regional innovation systems [1, 2]. One of the important attributes contributing to the S&T system performance is scientific collaboration [1, 6]. Scientific collaboration provides several salient advantages as shown in Table 2.1. One of the important factors leading to advantages of scientific collaboration is the social dimension of scientific work such as informal conversational exchanges between colleagues [8, 16], co-authorship relations [8, 19], jointly submitted grant proposals [8, 20], and co-patent applications [21-24].

Collaboration among scientists dates back to the 17th century [77], and it has become increasingly prevalent over the last two decades [7, 78]. Sonnenwald (2007) defined scientific collaboration as the interaction within a social context among two or more scientists in order to facilitate the completion of tasks with regard to a commonly shared goal. Thus, participants in the collaboration event integrate valuable knowledge from their respective domains to create new knowledge. According to the definition, collaborations must be perpetuated through social networks [51]. Therefore, social network analysis (SNA) is the method which is commonly used to reveal the structure of collaboration between researchers [6, 7, 27, 42, 43, 79, 80].

2.2.3. Relationship between Researchers' Communication and Their Collaborative Outputs

Co-authorship in scholarly publications is the most tangible and well-documented forms of scientific collaboration, and it is also a good indicator of the S&T system performance. Therefore, it is used widely in scientific collaboration studies [1, 8, 14, 19, 80, 81]. For example, using SNA, Newman (2001, 2001a, 2001b) and Barabási et al. (2002) analyzed the structural properties of scientific collaboration patterns in large scale by depicting the network of researchers when two authors were considered linked if their names appeared in the same scientific journal. They found that co-authorship networks were small world networks in which most nodes (i.e., authors) could be reached from other nodes by a small number of steps. With a similar approach used in co-authorship network studies, some studies also analyzed the structure of co-inventor maps in the case that two patent applicants (i.e. co-authors) were linked if there was a patent application together by these two applicants; thus, a network of co-invention was constructed. However, analyzing co-inventor maps was not used as widely as analyzing co-authorship maps [22]. In addition, for the networks constructed from researchers' jointly submitted grant proposals, there was not to my knowledge any study in the literature analyzing the properties of these networks, their relations with other concepts, and related implications.

Many scholars argue that co-authorship alone is insufficient as a measure of research collaboration. For example, Katz and Martin (1997) pointed out that many cases of collaboration did not result in co-authored publications; for example, when researchers worked closely together but decided to publish their results separately due to the fact that they came from different fields and desired to produce single-author papers in their own discipline. Their study concluded that measuring co-authorship was a partial indicator of research collaboration. Melin and Persson (1996) also asserted that co-authorship was only a rough indicator of collaboration, even though significant scientific collaboration leads to coauthored publications in most cases. The qualitative study of Laudel (2002) determined different types of collaborations that were classified according to the content of contribution made by collaborators. Then, a collaborator was rewarded with a co-authorship depending on the level of his/her contribution. The assumption that co-authorship and research collaboration are synonymous was criticized by several other scholars for the following reasons: listing co-authors for purely social reasons [8, 16, 30], listing co-authors simply by the virtue of providing material or performing a routine task [8, 16, 31], making the colleagues 'honorary co-authors' [8, 16, 32], and listing co-authors who did not even communicate with each other during research collaboration (e.g., many publications in physics and astrophysics include hundreds of authors) [33].

Researchers should communicate to formulate research questions that address either experimental or theoretical problems and to disseminate their results in order to obtain feedback.

Fox (1983) stated that communication and exchange of research findings and results were the most fundamental social process of science, and the principal means of this communication was the publication process. Communication between researchers not only stimulates them to think regarding the unsolved problems in their fields and possible research projects, thereby developing new ideas and solutions, but it also transmits ‘know-how’ or the procedural knowledge to efficiently solve the problems to other researchers [29]. Then, communication is an important source and influential factor for scientific collaboration [6, 8, 11, 41] and a fundamental component to sustain collaboration [7, 82]. Collaborations mostly begin informally and arise from informal communication between researchers, i.e., through close personal contacts and professional network [8, 16, 30, 34-36]. Since solving a scientific problem requires different complex tasks varying in uncertainty, much collaboration either do not occur or break up before becoming successful without informal communication [83]. Many scholars make a clear distinction between researchers’ communication and collaboration. For example, Melin and Persson (1996) reported that “collaboration was an intense form of interaction that allowed for effective communication”. Melin (2000) discussed that collaboration could be measured in a number of ways such as exchange of phone calls and e-mails, but a more concrete form to measure the collaboration was through co-authorship information. Laudel (2002) accepted publications as a way of formal communication, and found out that a considerable proportion of collaborations were not rewarded as a co-authorship. Borgman and Furner (2002) discussed that collaboration was one of the communication behaviors exhibited by authors in their various capacities. Similarly, from a network viewpoint, Newman (2001b) reported that there was an assumption that most people who wrote a paper together might not be genuinely acquainted with one another. Taking the assumption reported by Newman (2001b), one notable study made by

Pepe (2011) compared the structure of researchers' communication network with the structure of their collaborative output network (e.g., co-authorship network) by utilizing techniques used in SNA. The study found out the extent to which the structure of researchers' communication network overlaps the structure of their collaborative output network. That is, the more these network structures overlap the more likely collaborative output relations between researchers can be seen as a surrogate or proxy for communication relations between researchers. The study of Pepe (2011) can be extended into multiple networks constructed using the data collected via researchers' self-reports. In addition, Lee and Bozeman (2005) also reported the need of investigating whether collaboration structures of researchers really mimic communication structures of researchers. Based on the discussion so far, the following two hypotheses are proposed:

Hypothesis 1: Researchers' communication networks should highly overlap with their collaborative output networks.

Hypothesis 2: Researchers' communication networks positively impact their collaborative output networks.

2.2.4. Relationship between Researchers' Demographic Attributes and Their Collaborative Outputs

"Birds of a feather flock together" is the proverbial expression of homophily, which is often used in the social network literature. Macpherson et al. (2001) defined homophily as "the principle that a contact between similar people occurs at a higher rate than among dissimilar people". That is, it is more likely that individuals who share the same demographic attributes such as gender and race tend to interact with each other and to form social ties [84-86]. For example, Marsden (1988) found that individuals who shared the same race are more likely to discuss important matters. Similarly, Bozeman and Corley (2004) found that female researchers were more likely to collaborate with female researchers. Spatial proximity impacts the interaction between researchers and might increase or decrease the likelihood to collaborate. For example, the Kraut and Edigo (1988) found out that researchers in a close physical proximity tended to collaborate more due to the changes in three properties of informal communication: increasing the frequency of communication, increasing the quality of communication, and reducing the cost of communication. Olson and Olson (2000) also reported that face-to-face communication facilitates the flow of situated cognitive and social activities due to some of its key characteristics such as rapid feedback and multiple channels (e.g., voice, facial expression, gesture, body posture). However, the use of information and communication technologies (ICT) such as audio and video conferences, mobile phones, e-mail, social networking sites especially designed to support collaborative environment, and the World Wide Web facilitate informal communication between researchers and help them collaborate with other distant researchers in a timely manner [7, 39, 40]. Using both types of communication, face-to-face and ICT, have their own advantages and disadvantages [38]. Furthermore, given the possibility of disciplinary boundaries, the impact of a researcher's departmental affiliation should also be tested for each collaborative output network. To sum up, since collaboration is related to many types of shared attributes [16, 30], the aforementioned four networks should be analyzed by taking some demographic attributes of individuals such as gender, race, departmental affiliation, and spatial proximity into consideration. Then, the following hypothesis was tested:

Hypothesis 3: Researchers who share the same attributes are more likely to form a collaborative output tie than researchers who do not.

2.3. Method

2.3.1. Sample and Questionnaire

The University of South Florida's College of Engineering has researchers who hold both tenured and tenure-track faculty positions, research associates, visiting professors, and graduate students to run the research. This study surveyed the entire population, which was comprised of 107 researchers who hold both tenured and tenure-track faculty positions. Research associates, visiting professors, and graduate students were not considered in this study. The dean of the College of Engineering, 1 researcher who was on leave of absence during the data collection period, and 5 researchers who were recently hired, totaling 7 researchers, were excluded. Therefore, the sample size was reduced to 100 researchers. Table 2.2 shows the breakdown of the sample size in terms of demographic attributes. There are 6 departments in the College of Engineering: Chemical and Biomedical Engineering (CBE), Civil and Environmental Engineering (CEE), Computer Science and Engineering (CSE), Electrical Engineering (EE), Industrial and Management Systems Engineering (IMSE), and Mechanical Engineering (ME).

The questionnaire was in the paper-and-pencil format. It was first designed in a web format (http://orisurvey.eng.usf.edu/). However, several researchers during the pilot test or others later commented that filling out the questionnaire in a paper-and-pencil format was easier and more comfortable. Before distributing the questionnaire to all researchers, a researcher from each department was randomly chosen and contacted to conduct a pilot test for the questionnaire. Based on the comments and feedback from the researchers, the content and layout of the questionnaire were updated to facilitate gathering the responses. The questionnaire was 3 pages long and contained a total of 26 questions (see the Appendix A). The first page included 2 questions and respondents were asked to make a self-report of the number of both in-progress and completed collaborative outputs with other researchers with whom they engaged in co-authored or joint publications (in-preparation, [re]submitted or rejected, and published), joint grant proposals (in-preparation, declined, and funded), and joint patents (rejected, submitted, and issued) as well as researchers' names (see the Appendix A). The names of the researchers from 6 different departments within the college were already populated in 6 different tables in order to facilitate the thought process of the respondents. Each table had a different number of rows due to the different number of researchers in each department and 5 columns. The first 2 columns contained the last name and first name information of the researchers populated for each department. The third, fourth and fifth columns were the columns into which the respondent put the number of total in-progress and completed joint publications, grant proposals, and patents with other researchers. Since it might be hard for the respondents to remember the exact number of their total in-progress and completed collaborative outputs with other researchers, an ordinal scale was used to facilitate the thought process of the respondents. In the scale, the scores 1, 2, 3, and 4 were assigned to the number of collaborative outputs of 1 to 2, 3 to 5, 6 to 9, and 10-above, respectively. For example, if a respondent has either 1 or 2 joint publications with another researcher the respondent scans the names in the tables and puts the score 1 into the related cell next to the researcher's name under the publication column. If a respondent has 3, 4, or 5 joint grant proposals with another researcher the respondent finds the his/her collaborator's name in the tables and put the score 2 into the related cell next to the researcher's name under the grant proposal column. The respondents were also asked to provide their collaborators' names outside of the college and to put the number of in-progress and completed collaborative outputs with those collaborators at the bottom of the page. The second page included 4 questions and respondents were first asked to report the names of researchers with whom they exchanged conversations or ideas as well as the frequency of the exchange (see the Appendix A). A researcher's frequency of communication with other researchers and strength of closeness and intimacy in their communication ties with other researchers were assessed by a second, third and fourth question, and were rated based on a 6-point Likert-type scale, 6-point Likert-type scale, and 5-point Likert-type scale, respectively. These questions, denoted by Q2, Q3, and Q4, referring to three dimensions of tie strength in the social network literature. Tie strength can be assessed by three indicators: the frequency of conversational exchange (Q2), the intensity of the conversational exchange (Q3), mutual confiding or level of intimacy between conversational partners (Q4) [70, 71]. The second page was the same as the first page except that columns next to the columns across which the researchers' names were populated were kept for reporting the answers for the Q2, Q3, and Q4. Moreover, the respondent follows the same procedure which was followed to fill out the questionnaire on the first page. For example, a researcher scanned the names in the table, found his/her conversational partner's name, and put a score for the frequency of communication and the strength of closeness and intimacy into the cell next to the researcher's name in a given scale. The third page included the assessment of perceived innovativeness [57, 59, 87]. There were 20 questions each of which was marked in 5-point Likert scale (see Appendix A).

Information for the relations of both the communication (i.e., conversational exchange) and collaborative outputs between researchers was asked for the last 6 years up to current study date (between 2006 and 2012). This length of time might be reasonable for reporting the relations of the collaborative outputs, but not of communication because two researchers, for example, talk to each other frequently while they write a journal or proposal, but when they finish writing the journal or proposal they do not talk as frequently as they talked in the past.

However, the main point was to investigate to what extent the researchers were genuinely acquainted with one another on average from the self-perception perspective. In addition, the time frame, 6 years, must be the same to maintain a balanced comparison between networks constructed from the relations of both the communication and collaborative outputs.

2.3.2. Data Collection

The researchers were asked to complete a three-page questionnaire in three steps. First, a mass e-mail from the dean's office was sent out to the researchers in the sample, indicating that each of the researchers would be contacted through either their affiliated department or e-mail. Second, a graduate student from the college of engineering contacted the researchers by either joining their departmental meetings or e-mailing each researcher. The student handed out the paper-and-pencil questionnaire to each researcher in the meeting and made a short presentation about the details of the questionnaire. Additionally, the questionnaire was e-mailed to the researchers who were not present in the meetings as an attachment. Last, the graduate student followed up with each researcher in the sample in 2-3 weeks for completed questionnaires via e-mail. Completed questionnaires collected from the participants by visiting them directly to protect the confidentiality of their responses. If the questionnaire was not completed yet, an additional one week was given to the participants for completion before collecting the questionnaires directly from the participants.

Response rates were very low at the end because the number of both fully and partially completed questionnaires received was about 10. Therefore, to increase response rates, each researcher was also contacted personally both to make an in-person delivery of the questionnaire and to explain the purpose of the study and the details. The researchers were requested to fill out the questionnaire without using any forceful action which was against the protocol guidelines in the informed consent. Dillman (2007) discussed the factors improving response rate which can be achieved by in-person delivery. Two of those were observed in this study. First, a deliberate effort was made to increase the salience of the experience of receiving the questionnaire; thus, the interaction time required for presenting the questionnaire to the researcher was lengthened. Second, responsibility was assigned to a researcher rather than addressing the request in a general way.

Contacting the researchers personally was performed in two steps. First, the graduate student contacted the researchers personally to deliver the questionnaire in person, explained the details of the paper-and-pencil questionnaire face-to-face, and asked for whether they were willing to participate in the questionnaire or not. Later, the researchers who were willing to participate either filled out the questionnaire at the time they were contacted personally or made an appointment with the graduate student to fill later or filled on their own. The presence of graduate student was helpful because the researchers asked if they had any questions. The questionnaire was completed in 15-20 minutes on average; however, a few took more time to complete the questionnaire. A total of 76 out of 100 tenured/tenure-track faculty members participated in the questionnaire. Table 2.2 shows the breakdown of the participants in terms of demographic attributes. It took almost one semester to reach out to the target faculty members and to finalize all responses from the participants. Table 2.3 shows the timeline of the steps taken. One potential risk in this study was the low participation rate while collecting the social network data of researchers. If the participation rate is low, it is difficult to entirely depict connections between researchers, opening up the possibility that the results found in the analyses of the networks will be misleading. However, even if a particular faculty member did not fill out the questionnaire, the connections to non-participants are reported by the participants. Thus, connections of non-participants can be obtained from the perspective of participants. At the end, collaboration information for the full list of researchers is obtained. In this study, information about the connections of 24 non-participants was obtained by utilizing the best possible scenario explained in the next section. Another risk in this study was that the respondents might rate the Q2, Q3, and Q4 on the second page for all researchers because the respondents might think that they at least held a minimum relationship with any other researcher even if they did not communicate with them. For example, there were only two respondents who rated the Q2, Q3, and Q4 with all minimum scores for all other researchers within the sample. Therefore, for only these two respondents, the respondents' ratings for the Q2, Q3, and Q4 all of which received minimum scores for all other researchers were dropped while constructing the data matrixes.

2.3.3. Constructing Social Network Data Matrixes

This study focuses on the population of research faculty within the University of South Florida's College of Engineering. Data were collected by employing a questionnaire by which researchers report their contacts, the number of collaborative outputs, and the frequency of communication with them in a self-reported manner. The relational data obtained through the questionnaire was put into the form of a two-way matrix where rows and columns referred to researchers making up the pairs [49]. Furthermore, each cell in the matrix indicated the collaborative output or communication ties between the researchers. Thus, four 100x100 matrixes were constructed from the relational data provided by the researchers: a matrix of communication relations and a matrix of joint publications (or co-authorship), grant proposals, and patents. A total of 125 extra names were reported outside of the college through the name generator located at the bottom of the page in the questionnaire. However, these names were not included while constructing the matrixes in order to maintain the balanced comparisons between researchers' social network metrics (e.g., degree centrality) for further analyses and kept for a future study.

Five possible cases of reciprocity happened between two researchers when they rated each other regarding their connections:

1. Both researchers rated each other with an equal score for the frequency of communication and the number of collaborative outputs. In other words, the case was that the values of the upper and lower triangle cells were equal to each other in the 100x100 matrixes.
2. Both researchers rated each other with a different score for the frequency of communication and the number of collaborative outputs. In this situation, two cases might happen.
   a. One case was that the value of the upper triangle cells was *higher* than the value of the lower triangle cells in the 100x100 matrixes.
   b. The other was that the value of the lower triangle cells was also *higher* than the value of the upper triangle cells in the 100x100 matrixes.
3. Only one of the researchers rated the other. In this situation, two cases might also happen.
   a. One case was that the upper triangle cell contained a value, but lower triangle cell did not in the 100x100 matrixes.
   b. The other was that the lower triangle cell contained a value, but the upper triangle cell did not in the 100x100 matrixes.

Table 2.4 summarizes the five possible cases of reciprocity seen in the 100x100 matrixes when at least one researcher in a pair gives a non-zero rating to the other. 'X' and '0' indicate the ratings happening on only one side and non-ratings, respectively. Table 2.5 illustrates the number of occurrences of these cases in each network. The inter-rater agreement (IRA) percentage in a network was calculated by dividing the total number of occurrences in 'Equal-Equal' cases by the total number of occurrences of all cases (e.g., 120 was divided by 1234 which is the sum of 120, 141, 144, 377, and 452 for the network of communication). In IRA percentage calculation, the cases where both sides did not report a tie to the other (i.e., the cases where both sides score 0) were neglected. For the purpose of this study, directionality of the networks is not of fundamental importance [33]. This is because the collaborative output networks such as co-authorship networks are analyzed as undirected in the literature. Therefore, reported reciprocity in the number of collaborative outputs was converted to undirected edges. In order to make an equivalent comparison between the networks, the reported reciprocity in the frequency of communication was also converted to undirected edges. The researchers' social network data matrixes were symmetrized by converting the reported reciprocities to the undirected edges according to the most idealistic scenario shown in Table 2.6. In social network analysis, this symmetrization principle is known as the "maximum" method [89].

2.4. Results

Hypotheses in this chapter and next chapters were tested using SNA metrics and techniques. In order to both compute SNA metrics and perform SNA techniques, a computer package for SNA(UCINET version 6.308), a statistical computing software (the R project, called shortly 'R'), and a free and open network overview, discovery and exploration add-in for Excel 2007/2010 (the NodeXL) are used [89-92].

2.4.1. Visual Inspection of Networks

The NodeXL was used to visualize the networks. A graph is the mathematical structure that models a network with an undirected dichotomous (or binary) relations i.e., ties that are either present or absent between each pair of actors [49]. Graphs for four networks were depicted in Figure 2.1 using the 'Hare-Koren Fast Multiscale' layout option in which the isolated nodes are not shown. In each graph, a vertex (or node) refers to a researcher, and an edge refers to the relations of either communication or collaborative outputs between researchers. The network densities can be easily noticed from high to low as follows: network of communication, joint grant proposals, joint publications, and joint patents.

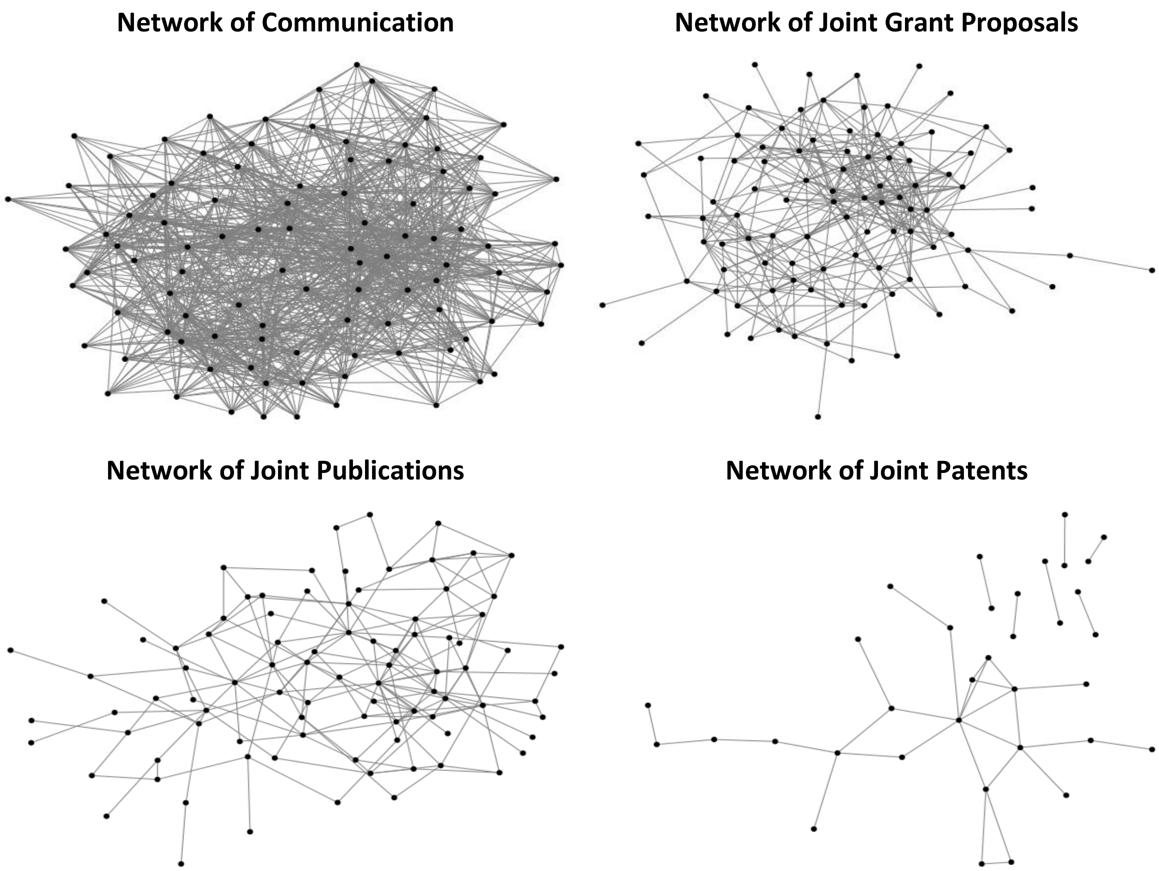

Figure 2.1 Visualization of Researchers' Communication and Collaborative Output Networks

2.4.2. Statistical and Descriptive Properties of Networks

Table 2.7 illustrates the different statistical and descriptive properties of four networks. *A connected component* of a graph is a maximal connected subgraph in which any two nodes are connected to each other by paths, and also there is no path between a node in the component and any node that is not in the component [49]. *Single-vertex connected component* in a graph is the isolated nodes, i.e., nodes which do not have any connections with other nodes. The network of communication, joint publications, and joint grant proposals had one connected component, while the network of joint patents had 7 connected components, meaning that there were 7 maximally connected subgraphs. The network of communication had no isolated nodes, whereas the network of joint publications, joint grant proposals, and especially joint patents had several isolated nodes.

Density of a graph, denoted by D, is the ratio of the number of edges present, L, to the maximum possible edges, $n(n-1)/2$, in a undirected graph, where n refers to the number of nodes [49]. That is, it is calculated as:

$$D = \frac{L}{n(n-1)/2} = \frac{2L}{n(n-1)}. \tag{2.1}$$

Density of a valued graph, denoted by D_v, is the sum of all valued edges, $L_{valued} = \sum_1^L V_L$, where L is the number of edges present and V_L is the value attached to an edge, divided by the maximum possible edges [89]. That is, it is calculated as:

$$D_v = \frac{L_{valued}}{n(n-1)/2} = \frac{\sum_1^L V_L}{n(n-1)/2}. \tag{2.2}$$

The result for network density for both binary and valued relations range from highest to lowest in the following order: communication, joint grant proposals, joint publications, and joint patents. Since the type of the rating scale used to construct the network of communication is different from other collaborative output networks, the valued density computed for the network of communication is much higher than the valued density computed for other collaborative output networks. The results indicate that the researchers' network relations generating the collaborative outputs are sparser than their network of communication relations.

A shortest path between two nodes is referred to geodesic. Geodesic distance or distance between two nodes is defined as the length of any shortest path between them, i.e., the number of edges connecting two vertices in a shortest path [49]. *Maximum geodesic distance or diameter* of a graph is the length of the largest geodesic distance any pair of nodes. The diameter of a graph quantifies how far apart two nodes are located in the graph [49]. If a graph is not connected, both distance and diameter are infinite or undefined because distance between some pairs of nodes is infinite in a disconnected graph [49, 93]. The NodeXL computes the diameter of the connected component and does not consider the isolated and disconnected subgraphs in the computation. Diameter of the connected component for the network of communication, joint publications, joint grant proposals, and joint patents is 3, 7, 7, and 9, respectively. This can be interpreted as that an idea can travel from any researcher to any other researcher over no greater than 3, 7, 7, and 9 steps. *Average geodesic distance* (AGD) is the sum of shortest paths between each vertex pairs divided by the number of possible vertex pairs, i.e., the average number of steps to connect any two nodes in a network [27, 33]. The number of possible vertex pairs is computed by n(n-1)/2 in a undirected graph, where n refers to the number of nodes. The AGD value, 1.792, is lower in the researchers' communication network than other networks: co-authored or joint publications, 3.468, joint grant proposals, 2.699, and joint patents, 3.452. This can be interpreted as that an idea can travel from any researcher to any other researcher over an average of 1.792, 3.468, 2.699, and 3.452 steps.

Clustering coefficient (CC) is defined as a measure of the extent to which nodes tend to cluster together in the network [93]. It can also be defined as "the average fraction of pairs of a person's collaborators who have also collaborated with one another" [27]. Clustering coefficient for whole network, CC, is found by averaging the local clustering coefficients of all vertices n [94]. Local clustering coefficient, LCC of vertex i from vertices n, is computed by dividing the number of edges among the neighbors of vertex i by maximum possible edges of the neighbors of vertex i [94]. Clustering coefficient for whole network, CC, is found by averaging the local clustering coefficients of all vertices n [94]. Both of them are calculated as:

$$LCC_i = \frac{\text{Number of edges among neighbors of vertex } i}{\text{Maximum possible edges of neighbors of vertex } i} \tag{2.3}$$

$$CC = \frac{1}{n}\sum_{i=1}^{n} LCC_i . \tag{2.4}$$

As seen from the formula, the LLC calculates the density of an ego's neighbors, but by leaving out the ego [93]. In other words, it computes the density of connections among nodes that are already connected through two-path. The CC value, 0.534, is higher in the researchers' communication network than other networks: joint publications, 0.158, joint grant proposal, 0.285, and joint patent, 0.051. Then, the results indicate that two researchers have a 53.4% chance of communicating and a 15.8%, 28.5%, and 5.1% chance of collaborating in publications, grant proposals, and patents, respectively if they have both communicated and collaborated with another third researcher. In other words, for researchers' communication relations, two individual researchers have 53.4% chance of being acquainted with one another through a common researcher who puts them in contact in the College of Engineering. For researchers' joint publication, joint grant proposal, and joint patent relations, two individual researchers have 15.8%, 28.5%, and 5.1% chance of being acquainted with one another through a common researcher who puts them in contact in the College of Engineering, respectively. This means that collaborations in a group of three or more researchers for grant proposals are more common than collaborations in a group of three or more researchers for publications and patents. A small-world network is a network in which most nodes can be reached by any other in a small number of steps [95]. Two properties are observed in the small-world networks: 1) higher clustering that it would be expected by chance 2) AGD on average are as short as it would be expected by chance. Many real networks have the property of being a small-world in which AGD is low, while CC is high [94]. Then, network small-worldliness can be decided by comparing CC and AGD of a given network to a distribution of CC and AGD that were obtained from randomly generated graphs with an equivalent density and the same degree distribution. Since CC is another density measure, but for pairs that are already connected indirectly, the density of an original graph can be used as a rough sort of gauge for what you expect CC to be by chance. The binary graph densities in the networks of communication, joint publications, joint grant proposals, and joint patents are almost half, one-fourth, one-fourth, and one-fifth the value of their CCs, respectively. Therefore, there is a sense in which there is actually much more clustering than you do expect by chance in the collaborative output networks than there is in the communication network. One reason for this can be that specialization of the researchers in different areas of focus not only helps the formation of dense clusters of researchers but also encourages them to form short connections to other researchers. That is, specialization of the researchers helps them detect a common point of view for their research in conversations and brings them together for further collaboration. Still, the other property that is getting AGD as short as it would be gotten in a random graph must be tested to be able to fully decide on whether or not the networks shows a small-world property.

Assortativity (degree) is a measure of extent to which nodes with similar degree centralities tend to attach to one another, i.e., it is the measure of correlation in the degrees of connected nodes [96, 97]. There is a hypothesis that positive assortativity is a property of many socially generated networks, while negative assortativity is more prevalent in technological and biological networks [95]. Assortativity that is greater than 0 indicates that prolific authors tend to be connected with only prolific researchers. Assortativity that is less than 0 indicates that prolific authors tend to be connected with both prolific and non-prolific researchers [33]. The value of assortativity, -0.146, is less than 0 in the researchers' communication network, while it is higher than 0 for the other networks: joint publications, 0.044, joint grant proposal, 0.072, and joint patent, 0.217. The fact that the researchers' communication network has a negative assortativity means that when a newcomer is introduced into the network; the newcomer will not feel himself/herself a stranger to others and will begin to collaborate in an inclusive environment. On the contrary, a newcomer will tend to produce collaborative outputs with only prolific researchers in other networks: joint publications, grant proposal, and patents.

Distance-based cohesion, i.e., compactness, is calculated by harmonic mean of entries in the distance matrix and measures the degree of cohesiveness in the network from the distance perspective [89, 98]. The value ranges from 0 (nodes are completely isolated) to 1 (each node is adjacent, making up of a clique of all nodes). *Distance-weighted fragmentation* is 1 minus distance-based cohesion. The highest cohesive network with value of 0.618 is the researchers' network of communication, which is expected because every researcher is expected to communicate each other within the college. The second highest cohesive network with value of 0.401 is the researchers' network of grant proposals. Third and the last one is the researchers' network of joint publications and joint patents with the value of 0.277 and 0.021, respectively.

Number of conversational partners or collaborators per researcher is the ratio of the number of researchers' total conversational partners or collaborators to the total number of researchers. The number of researchers' total conversational partners or collaborators is computed by summing the upper or lower triangle rows in the data matrixes constructed according to Table 2.6 and the total number of researchers is 100. Then, for the network of communication, joint publications, joint grant proposals, and joint patents, the ratio is calculated as 12.34 (highest), 1.96, 3.67, and 0.35 (lowest), respectively. As seen from the results, the ratio for joint grant proposals is twice as much as the ratio for joint publications.

Table 2.8 illustrates the comparison of the density of four networks. In other words, it shows the degree to which the density of one type of relation among researchers is different from the density of another type of relation among the same researchers [93]. The conventional approach of calculating the standard errors assumes independent observations. However, using the conventional approach in the network data can be misleading because the conventional approach underestimates the true sampling variability due to dependency of the observations. Therefore, it gives too optimistic results due to underestimated sampling variability and leads to reject the null hypothesis that two densities are the same [93]. Using 'bootstrapping', a non-parametric sampling technique, a sampling distribution of densities of two networks is constructed. Standard deviations (called standard error) of these two the sampling distributions is used to calculate t-statistic when comparing the densities of two networks [99]. Thus, independence of observations is considered by accounting for the variation from sample to sample just by random chance [93]. When the ties are binary in the compared networks, the test is for a difference in the probability of a tie of one type and the probability of a tie of another type [93]. When the ties are valued in the compared networks, the test is for a difference in the mean tie strengths of the two relations [93]. The standard deviation and mean differences are illustrated in Table 2.8, which are obtained by both the classical method and the bootstrap sampling method, for both binary and valued relations. Comparison of the valued relations of the network of communication to other networks was discarded to maintain the balanced comparison because the type of the rating scale used to construct the network of communication was different from the type of the rating scale used to construct the collaborative output networks. It is noted that the mean difference by the classical method is almost the same as the mean difference by the bootstrap sampling method, while the standard deviation difference by the classical method is always smaller (i.e, underestimated) than the standard deviation difference by the bootstrap sampling method. The difference between densities for all network pairs is statistically significant. In other words, the observed difference would rarely be seen by chance in random samples drawn from these networks.

2.4.3. Network Comparisons

Correlation between two networks is computed by using the quadratic assignment procedure (QAP) technique. Since observations in dyadic data is interdependent, the traditional OLS technique to test the significance of the correlation between two networks cannot be used [100]. Therefore, an alternative technique, QAP, was first suggested by the statistician Mantel (1967), and it was later used by Hubert (1987) in a vast array of applications [103]. The procedure works in two steps. First, it computes the correlation coefficient between corresponding cells of the two data matrixes. Later, it randomly and synchronously permutes the rows and columns of one matrix and recomputes the correlation [89]. The second step is performed thousands of times in order to calculate the proportion of times that a random measure is higher than or equal to the observed measure calculated in the first step. A low proportion when compared to the desired significance level suggests that there is a strong relationship, which is unlikely to be occurred by a chance, between the matrixes, i.e., the correlation between two networks are statistically significant [89]. The Jaccard coefficient and Pearson correlation can be used to evaluate binary and valued relations, respectively [93]. Table 2.9 illustrates the QAP correlation results for both binary and valued relations. By QAP correlation results, *hypothesis 1* tests the extent to which researchers' communication network overlap with multiple collaborative output networks. All pairs of correlations are positive and statistically significant in both binary and valued relations, but the overlap of the researchers' communication ties with their collaborative output ties is not high. This shows that the acquaintanceship between researchers is not sufficiently reflected on joint collaborative output relations. The networks of joint publications and grant proposals are highly correlated, which is expected because grant proposals are generally written with the intention of publishing the results. In binary relations, the correlation between the network of communication and joint publications is *lower* than the correlation between the network of communication and joint grant proposals. At binary level, this implies that the idea exchanges between researchers that result in joint publications is not as common as the idea exchanges between researchers that resulted in joint grant proposals. Similarly, in valued relations, the correlation between the network of communication and joint publications is also *lower* than the correlation between the network of communication and joint grant proposals. At valued level, this implies that the correlation between the frequency of communication and the number of joint publications is lower than the correlation between the frequency of communication and the number of joint grant proposals. Among the correlations of the network of joint patents to other networks, the highest correlation is the one with the network of joint publications at both binary and valued level. This implies that there is a tendency among researchers that their joint publications were turned into joint patents in a collaborative manner.

QAP technique was also run to test whether or not researchers who have a similar spatial proximity tend to communicate more and produce more collaborative outputs together. For this, a spatial proximity 100x100 matrix, W, in which rows and columns refer to researchers making up the pairs, was constructed. The (i,j) element of W matrix, denoted $w_{ij,}$ quantifies whether or not two researchers are in the same neighborhood; in other words, the w_{ij} defines neighborhood structure over an area [104]. In this study, the case of whether or not two researchers are in the same neighborhood was measured on a scale: (1) different buildings, (2) the same building, (3) the same hallway, (4) next to each other. First, an *upper triangular* spatial proximity matrix was constructed. Later, it was symmetrized in order to obtain the 100x100 matrix. Table 2.10 illustrates the QAP correlation results between spatial proximity matrix and each collaborative output matrix (i.e., each collaborative output network) for valued relations. All pairs of correlations were positive and statistically significant.

2.4.4. Network Prediction

QAP technique can be used for regressing one network (dependent variable) on other networks (independent variables). Krackhardt (1988) first showed that beta parameters in an ordinary least squares-OLS model of network data could be tested using a multiple regression extension of QAP technique, MRQAP [103]. QAP first performs an OLS regression in order to estimate the regression coefficients (i.e., original regression coefficients) on the original dependent variable matrix. Second, the rows and columns of dependent variable matrix are randomly and synchronously permuted to obtain a mixed-up matrix, and another OLS regression is run for obtaining the new regression coefficients using this newly permuted dependent variable matrix. This procedure is done several times (in this study, 10000) to find the large set of OLS regression coefficients using a new randomly permuted dependent variable matrix at each time. The regression coefficients and R^2 are stored away after running each regression. Finally, the original regression coefficients are compared against the distribution of the stored regression coefficients and R^2's, which are obtained under the set of permuted regressions, for each of the independent variables. If fewer than 5% of the regression coefficients (i.e., betas) are larger than the observed regression coefficient, then the coefficient is considered significant at the 0.05 level, and the same is valid for the 0.01 level of significance [103, 106]. In this study, the Double Dekker Semi-Partialling MRQAP procedure was used since it gives more robust results. Unlike Y-permutation procedure, this procedure takes the correlation between independent variables into account by putting the resulting residuals, which are obtained from the regression of independent variables on each other, into the original regression equation [103].

Table 2.11 and 2.12 illustrates the regression with QAP technique for both binary and valued relations, respectively. For binary relations, researchers' communication relations had a positive and statistically significant impact on researchers' collaborative output relations. This impact was very minimal on researchers' joint patent relations. This implied that the communication between researchers had a positive impact on their collaborative outputs. While the impact of researchers' joint publication relations was high on their joint patents relations, the impact of researchers' joint grant proposal relations was low on their joint patents relations. This implied that joint publications between researchers were more likely to result in joint patents than joint grant proposals between researchers were. Additionally, the impact of researchers' joint grant proposal and publication relations on each other was high and statistically significant. This indicated that grant proposals was written by researchers in order to be able to get them published at the end. For valued relations, the network of researchers' communication relations had a positive and statistically significant impact on joint grant proposal relations. However, this impact became low and statistically significant on joint publication relations, and even negative and statistically significant on joint patent relations. This implied that the intensity in the frequency of communication between researchers resulted in only generating a greater number of joint grant proposals between them. The rest of the impacts were the same as discussed for binary relations.

Exponential Random Graph Models (ERGMs) (also called p^* models) can also be used to model the probability of observing a graph y from a random set of relations (edges and non-edges) Y using the various local (or subgraph) configurations, such as edges, triangles, reciprocated ties, k-stars, and etc., as independent variables expressed by the model [107-112]. In other words, the probability of observing a graph y depends on the presence of various configurations used as independent variables in ERGM model [108]. The distribution of Y can be parameterized in the following form:

$$P(Y = y \mid X) = \frac{\exp\left\{\theta^{\mathrm{T}} g(y, X)\right\}}{\kappa(\theta, Y)} \tag{2.5}$$

and, the above equation is the general form of the class of ERGM, where Y is the (random) set of relations (edges and non-edges) in a network, y is a particular given set of relations, X is a covariate (or a matrix of attributes) for the vertices and edges, θ is the vector of coefficients corresponding to a set of various type of configurations, $g(y, X)$ is a vector of network statistics corresponding to the related configuration included in the model if the configuration is observed in the network y, $g(y) = 1$; otherwise it is 0, and $\kappa(\theta, Y)$ is a normalization constant to let the probabilities sum to 1, and it is calculated as $\sum_{z \in Y} \exp\left\{\theta^{\mathrm{T}} g(z, X)\right\}$[107-111].

The above log-linear model can be turned into a logit model in the following form:

$$\log\left\{\frac{P(Y_{ij} = 1 \mid Y_{ij}^{C} X)}{P(Y_{ij} = 0 \mid Y_{ij}^{C} X)}\right\} = \theta^{T} \delta(y, X)_{ij}, \tag{2.6}$$

where $\delta(y, X)_{ij} = g(y_{ij}^{+} \mid X) - g(y_{ij}^{-} \mid X)$ is the vector of change statistics and y_{ij}^{+} and y_{ij}^{-} are the graphs where a tie from node i to node j is forced to be present (with y_{ij} =1) and absent (with y_{ij} =0), respectively, while all the rest of the network is exactly kept as in y itself [110, 113]. Y_{ij}^{C}

represents the rest of the network other than the single variable Y_{ij} [110]. Then, change in the network statistics $g(y, X)$ occurs when the tie from node *i* to node *j* changes from being present to absent [113]. Then, each coefficient θ can be interpreted as the increase in the conditional log-odds of network per unit increase in the network statistics $g(y, X)$ due to switch a particular Y_{ij} from 0 to 1 holding the rest of the network fixed at Y_{ij}^{C} [110].

The ties in the network of communication can be modeled as an edge covariate (i.e., independent variable) that affects the probability of the tie in the network of joint publications, grant proposals, and patents. Five separate models from the simplest to more complex were run taking the binary networks of the network of joint publications, joint grant proposals, and joint patents as dependent networks. Table 2.13 illustrates the results for all models. The package "ergm" in the R project for statistical computing was used to run the models [111]. *Model 1* is the simplest model that counts the equal probability for all edges in the network, and it is naturally null model from which to proceed and known as the Bernoulli model or the Erdős–Rényi model [109]. Then, *Model 1* can be shown as:

$$P(Y = y) = \frac{\exp\left\{\theta L(y)\right\}}{\kappa(\theta, Y)}, \tag{2.7}$$

where θ is the edge parameter and $L(y)$ refers to the number of edges in the graph *y* [107, 108]. The following models build up on *Model 1*. *Model 2* is to investigate whether or not the impact of the ties in the network of communication can influence the probability of ties in the network joint publications, grant proposals, and patents. Then, *Model 2* can be shown as:

$$P(Y = y) = \frac{\exp\left\{\theta L(y) + \sigma C(z)\right\}}{\kappa(\theta, Y)}, \tag{2.8}$$

where σ and $C(z)$ refer to the edge parameter for the network of communication and the strength of edges associated with the network of communication which is a graph *z*, respectively. Attribute information can be incorporated into an ERGM [110]. *Model 3* only considers the researchers' demographic attributes such as gender (0="female", 1="male"), race (1="Asian", 2="Black", 3="Hispanic", 4="White"), department affiliation (1="CBE", 2="CEE", 3="CSE", 4="EE", 5="IMSE", 6="ME"), and spatial proximity. Four 100X100 spatial proximity matrixes in which rows and columns refer to researchers making up the pairs, was constructed. The (*i*,*j*) element of each matrixes are dummy coded (as 1 and 0 otherwise) whether or not two researchers' offices are next to each other, and located in the same hallway, in the same building, and in different buildings. A dummy coded matrix indicating that researchers are located in the separate buildings was chosen as base proximity matrix. Then, the effect of the first three proximity matrixes is evaluated relative to the effect of the base proximity matrix. Then, *Model 3* can be shown as:

$$P(Y = y) = \frac{\exp\left\{\theta L(y) + r^T A(y)\right\}}{\kappa(\theta, Y)}, \tag{2.9}$$

where r^T is the vector of parameters for attributes (or covariates). While $A(y)$ is the vector of edge level covariates which refer to uniform homophily effect (i.e., individuals who share the same attribute are more likely to form social ties than two actors who do not share) for each attribute: gender, race, department affiliation, and spatial proximity. Unlike *Model 3*, *Model 4* includes σ and $C(z)$. Then, *Model 4* can be shown as:

$$P(Y = y) = \frac{\exp\left\{\theta L(y) + \sigma C(z) + r^T A(y)\right\}}{\kappa(\theta, Y)}, \tag{2.10}$$

Model 5 includes o_k and $O(t_k)$ which are the edge parameter for two collaborative output networks other than the collaborative output network modeled as dependent variable and the strength of edges associated with these networks, respectively. Then, *Model 5* can be shown as*:*

$$P(Y = y) = \frac{\exp\left\{\theta\ L\ (y) + \sigma\ C(z) + r^T A(y) + o_k O(t_k)\right\}}{\kappa(\theta, Y)}, \tag{2.11}$$

where *k*=1 and 2 referring to a couple of collaborative output networks other than the collaborative output network modeled as dependent variable. The results of these models are discussed below.

Model 1 is the 'edges' term that acts as the 'intercept' for the model. It is based on the number of edges (or density) of the observed network (compare the density of the networks in Table 2.7 for binary relations with ERGM results). ERGM fits a type of logistic model, so to interpret the parameter estimate; one must use the logistic transform because the coefficients are expressed as conditional log-odds. The value of -2.525 (log odds) means that the addition of any edges to the network of joint grant proposals changes the total number of edges by the probability of 0.074 (calculated by e-2.525/(1+e-2.525)). In other words, the probability that a tie that is completely heterogeneous will form in the network of joint grant proposals is 0.074. The probabilities that a tie that is completely heterogeneous will form in the network of joint publications and joint patents are 0.040 and 0.007, respectively.

Model 2 tests the impact of researchers' communication network ties on their collaborative output ties without considering any other demographic attributes. The probability of a tie in the network of joint grant proposals is increased by a log-odds factor of -3.753+0.730*(*n*) for every unit increase in the frequency score *n* in the network of communication. If the communication network score is the minimum 'once every three months

(1)’, this means that the addition of any edges with the value of strength ‘1’ to the communication network changes the total number of edges in the network of joint grant proposals by the probability of 0.046 (calculated by $e^{-3.753+0.730(1)}/(1+ e^{-3.753+0.730(1)})$), and if communication network score is the maximum ‘once a day (6)’, the probability of a tie in the network of joint grant proposals is 0.652 (calculated by $e^{-3.753+0.730(6)}/(1+ e^{-3.753+0.730(6)})$). Similarly, for the minimum and maximum communication network scores, the probability of a tie in the network of joint publications was 0.019 and 0.465, respectively, and the probability of a tie in the network of joint patents was 0.003 and 0.101, respectively. The results indicated that the probability of a tie that would form in the network of joint grant proposals was greater than the probability of a tie that would form in the network of joint publications and joint patents for the minimum and maximum communication network scores. These findings are similar to the findings was found by QAP regression that was run for valued relations.

Model 3 considers demographic attributes in addition to the ‘edge’ parameter. For joint grant proposals as the dependent network, the log-odds of a tie that is homogenous by either race only, department only, and closer than being in different buildings only are -2.860 (=-3.212+0.352), -1.888 (=-3.212+1.324), -2.404 (=-3.212+0.808) – being next to each other, -2.517 (=-3.212+0.695) – being on the same hall, respectively. The attribute ‘gender’ is excluded because it is not statistically significant, and the attribute ‘being in the same building’ is also excluded for the same reason. Then, the corresponding probabilities that a tie which is homogenous by either race only, department only, and closer than being in different buildings only will form in the network of joint grant proposals are 0.054 (calculated by $e^{-2.860}/(1+e^{-2.860})$), 0.131 (calculated by $e^{-1.888}/(1+e^{-1.888})$), 0.083 (calculated by $e^{-2.404}/(1+e^{-2.404})$), and 0.075 (calculated by $e^{-2.517}/(1+e^{-2.517})$), respectively. The log-odds of a tie that is homogenous by race, department, and being next to each other is -0.728 (=-3.212+0.352+1.324+0.808) and the log-odds odds of a tie that is homogenous by race, department, and on the same hall is -0.841 (=-3.212+0.352+1.324+0.695). The corresponding probabilities are 0.326 (calculated by $e^{-0.728}/(1+e^{-0.728})$) and 0.301 (calculated by $e^{-0.841}/(1+e^{-0.841})$).

For joint publications as the dependent network, the log-odds of a tie that is homogenous by either gender only, race only, department only, and closer than being in different buildings only are -4.106 (=-4.487 +0.381), -4.010(=-4.487+0.477), -2.768(=-4.487+1.719), and -2.454(=-4.487+0.758) – being on the same hall, respectively. Attributes 'being next to each other' and 'being in the same building' are excluded because they are not statistically significant. Then, the corresponding probabilities that a tie which is homogenous by either gender only, race only, department only, and closer than being in different buildings only will form in the network of joint publications are 0.016, 0.018, 0.059, and 0.023, respectively. The log-odds of a tie that is homogenous by gender, race, department, and on the same hall is -1.152 (=-4.487+0.381+0.477+1.719+0.758) which generates the corresponding probability as 0.240.

For joint patents as the dependent network, the log-odds of a tie that is homogenous by department only and closer to each other than being in different buildings are -4.991 (=-6.610 +1.619), -5.284(=-6.610 +1.326) – being next to each other, -5.948(=-6.610 +0.662) – being on the same hall, and -5.699 (=-6.610 +0.911) – being in the same building, respectively. Attributes 'gender' and 'race' are excluded because they are not statistically significant. Then, the corresponding probabilities that a tie which is homogenous by either department only and closer than being in different buildings only will form in the network of joint patents are 0.007, 0.005, 0.003, and 0.003, respectively. The log-odds of a tie that is homogenous by department and being next to each other is -3.665(=-6.610+1.619+1.326), the log-odds of a tie that is homogenous by department and on the same hall is -4.329(=-6. 610+1.619+0.662), and the log-odds of a tie that is homogenous by department and being in the same building is -4.080(=-6.610+1.619+0.911). The corresponding probabilities are 0.025, 0.013, and 0.017, respectively. The results indicated that the likelihood of the presence of a tie which is homogenous in all significant attributes was close to each other in the network of joint grant proposals and joint publications and it was the lowest in the network of joint patents.

Model 4 tests the impact of researchers' communication network ties on their collaborative output ties in the presence of all other demographic attribute effects. For the minimum and maximum communication network scores, the probability of a tie in the network of joint grant proposals was 0.060 and 0.811, respectively, while the other variables are held constant in the model. The effect of attribute 'being on the same hall' is not statistically significant, so it is excluded. Then, the probabilities of a tie that is homogenous by gender, race, department, being next to each other, and being in the same building are 0.014 for the minimum communication network score and 0.490 for the maximum communication network score.

For the minimum and maximum communication network scores, the probability of a tie in the network of joint publications was 0.015 and 0.428, respectively, while the other variables are held constant in the model. The effect of attributes 'gender', 'department', and 'being on the same hall', and 'being in the same building' is not statistically significant, so they are not considered. Then, the probabilities of a tie that is homogenous by race and being next to each other are 0.008 for the minimum communication network score and 0.275 for the maximum communication network score.

For the minimum and maximum communication network scores, the probability of a tie in the network of joint patents was 0.002 and 0.063, respectively, while the other variables are held constant in the model. The only statistically significant attribute is ‘being in the same building’. Then, the probabilities of a tie that is homogenous by being in the same building are 0.003 for the minimum communication network score and 0.120 for the maximum communication network score. When the results are compared with the probabilities obtained in *Model 2*, the likelihood of the presence of a tie which is homogenous in all significant attributes was decreased in the network of joint grant proposals and joint publications, whereas it was increased in the network of joint patents for the minimum and maximum communication network scores. In other words, when communication between researchers who shared the same attribute at both the minimum and maximum level was considered the effect that the ties in the network of communication would increase the likelihood of the presence of ties in the network of joint grant proposals and joint publications was diminished, whereas that effect was increased in the network of joint patents.

Unlike *Model 4*, *Model 5* considers other collaborative output network effects other than the network used as the dependent variable. There is positive and statistically significant log-odds effect of the network of joint publications on the network of joint grant proposals, whereas the log-odds effect of the network of joint patents on the network of joint grant proposals is not statistically significant. For the minimum and maximum communication network scores, the probability of a tie in the network of joint grant proposals was 0.056 and 0.720, respectively, while the other variables are held constant in the model. The effect of attribute ‘race’, ‘being next to each other’ and ‘being on the same hall’ is not statistically significant, therefore they are excluded. Then, the probabilities of a tie that is homogenous by gender, department, and being in the same building are 0.016 for the minimum communication network score and 0.412 for the maximum communication network score. The log-odds of a tie that is homogenous by gender, department, and being in the same building is -2.452(=-3.573+0.753*(1)-0.381-0.644+1.670*(1)) for the minimum communication network and collaborative output network scores and 6.323(=-3.573+0.753*(6)-0.381-0.644+1.670*(4)) for the maximum communication network and collaborative output network scores. The corresponding probabilities are 0.079 and 0.998, respectively.

There is positive and statistically significant log-odds effect of the network of both joint grant proposals and joint patents on the network of joint publications. For the minimum and maximum communication network scores, the probability of a tie in the network of joint publications was 0.007 and 0.074, respectively, while the other variables are held constant in the model. The effect of attribute ‘gender’, ‘department’, ‘being on the same hall’, and ‘being in the same building’ is not statistically significant, therefore they are not considered. Then, the probabilities of a tie that is homogenous by race and being next to each other are 0.005 for the minimum communication network score and 0.051 for the maximum communication network score. The log-odds of a tie that is homogenous by race and being next to each other is -0.357(=-5.376+0.474*(1)+0.519-0.920+1.549*(1)+3.397*(1)) for the minimum communication network and collaborative output network scores and 16.851(=-5.376+0.474*(6)+0.519-0.920+1.549*(4)+3.397*(4)) for the maximum communication network and collaborative output network scores. The corresponding probabilities are 0.412 and 1, respectively.

There is positive and statistically significant log-odds effect of the network of both joint grant proposals and joint publications on the network of joint patents. For the minimum and maximum communication network scores, the probability of a tie in the network of joint patents was 0.001 and 0.003, respectively, while the other variables are held constant in the model. None of attributes are statistically significant except ‘being in the same building’. Then, the probabilities of a tie that is homogenous by being in the same building are 0.003 for the minimum communication network score and 0.009 for the maximum communication network score. The log-odds of a tie that is homogenous by being in the same building is -4.114(=-7.142+0.224*(1)+1.112+0.569*(1)+1.123*(1)) for the minimum communication network and collaborative output network scores and 2.082(=-7.142+0.224*(6)+1.112+0.569*(4)+1.123*(4)) for the maximum communication network and collaborative output network scores. The corresponding probabilities are 0.016 and 0.889, respectively. Then, when the results are compared with the probabilities obtained in *Model 2*, unlike *Model 4* results, the likelihood of the presence of a tie which is homogenous in all significant attributes was decreased in each collaborative output network. Furthermore, it was observed that the likelihood of the presence of a tie in researchers' collaborative output networks is increased after including the effect of other collaborative output ties. Especially, this increase is drastic when the strength of other collaborative output ties are the maximum.

Hypothesis 2 tests whether or not the ties in the network of communication would increase the likelihood of the presence of ties in each collaborative output network. The following results were observed from *Models 2*, *4*, and *5*, when keeping other variables constant in *Models 4* and *5*. The ties in the network of communication significantly and positively impacted the likelihood of the presence of ties in each collaborative output network. The probability of a tie in the network of joint grant proposals was always higher than the probability of a tie in the network of joint publications and joint patents. For the minimum communication network score, the probability of a tie in all collaborative output networks almost remained at the same level. However, for the maximum communication network score, the probability of a tie in the network of joint grant proposals was increased, while the probability of a tie in the network of joint publications and joint patents was decreased as progressed from *Model 2* to *Model 5*. *Hypothesis 3* tests mainly the homophily hypothesis in which researchers who share the same attribute tend to form social ties more than researchers who do not. When *Model 5s* with the lowest AIC scores are the models chosen as the base models, the following are observed. Being of the same gender had a statistically significant negative effect on the network of joint grant proposals, whereas the effect of gender on the network of publications was not statistically significant. The results indicate that grant proposals are submitted with mixed gender teams in the college of engineering. That is, researchers perceive that their projects have a better chance to be funded if they have a gender diverse team. Sharing the same race attribute had a statistically significant positive effect on the network of both joint publications, whereas the effect of sharing the same race on the network of joint grant proposals was not statistically significant. This shows that the same race researchers are more likely to publish together. In other words, sharing the same race increases the chance of joint publications [114], whereas sharing the same race does not impact the chance of joint grant proposals. Being in the same department had a statistically significant negative effect on the network of joint grant proposals, but had no effect on the network of joint publications, indicating that there is a tendency of interdepartmental collaboration among researchers in joint grant proposals; however, whether or not researchers are affiliated with the same department makes no difference in their joint publications. Then, it can be said that grant proposal writing bridges departments to a much greater degree than publication does. Additionally, there was no effect of the demographic attributes ‘gender’, ‘race’ and ‘department’ on the network of joint patents.

The effect of being in the same level of spatial proximity varies for each collaborative output network. Being in the same building had a statistically significant negative effect on the network of joint grant proposals, meaning that researchers who are in the same building are less likely to collaborate for grant proposals compared with researchers who are in different buildings. Being next to each other had a statistically significant negative effect on the network of joint publications, indicating that the likelihood of researchers who are next to each other to collaborate for publications is less than the researchers who are in different buildings. Being in the same building had a statistically positive effect on the network of joint patents. This shows that being in the same building increases the likelihood of collaboration for patents compared with being in different buildings. To summarize, being closer to each other decreases the likelihood of collaboration for publications and grant proposals, but it increases the likelihood of collaboration for patents compared with being in different buildings. The more researchers are distant to each other the more likely they collaborate for publications and grant proposals [7, 39, 40]. For example, if this study was conducted to map interdisciplinary relations on a campus, the results would be highly expected that researchers from different colleges were more likely to form collaborative ties. Then, investment for an online collaborative website for researchers will be helpful to connect distant researchers to generate more collaborative outputs between them. Furthermore, research centers in which researchers are more spatially collocated will help increase the likelihood of formation of co-inventor relations.

The effect of the network of joint publications and grant proposals on each other was positive and statistically significant. Similarly, the effect of the network of joint patents and publications on each other was positive and statistically significant. These results match up with the QAP results. However, the effect of the network of joint patents on the network of joint grant proposals was not statistically significant, whereas the effect of the network of joint grant proposals on the network of joint patents was positive and statistically significant. This might be due to a temporal order of collaborative outputs. For example, researchers first start writing grant proposals to both obtain a publishable output and issue patent at the end. Also, joint publications and joint patents mostly occur simultaneously. Then, the case that the network of joint patents impacts the network of joint grant proposals becomes against the natural progression of collaborative outputs.

2.4.5. Centrality Comparisons

For all networks using, four types of normalized centrality metrics for each researcher, network centralization, and group degree centralities were computed, and hypothesis tests about mean centrality of groups were also performed. All centrality metrics were calculated using binary relations. These centrality metrics are as follows:

Degree Centrality of a node n_i, denoted by $C_D(n_i)$, is the number of nodes that adjacent to node n_i or the number of unique edges, e_{ij}, that are connected to node n_i [49]. Normalized degree centrality, $C'_D(n_i)$, is found by dividing the degree centrality of node n_i by the number of total nodes, n, excluding n_i such as (n-1).Then, Normalized degree centrality can be used to compare the degree centrality of nodes across networks of different size. Thus, $C'_D(n_i)$ which ranges from 0 to 1 is given by:

$$C'_D(n_i) = \frac{C_D(n_i)}{n-1} = \frac{\sum_j e_{ij}}{n-1}, \tag{2.12}$$

where $\sum_j e_{ij} = \sum_i e_{ji}$ for undirected networks.

Closeness Centrality of a node n_i, denoted by $C_C(n_i)$, is the sum of geodesic distances (i.e., geodesics) to all other nodes in a network [49]. Geodesic distance is a shortest path (i.e., lowest total number of edges) linking node, n_i and n_j, which is denoted by d(n_i, n_j). Then, the sum of geodesic distances is shown by $\sum_{j}^{n} \mathrm{d}(n_i, n_j)$. A lower closeness centrality score indicates a more central position for a node in a network [90]. Sabidussi's (1966) index of actor closeness offers the sum of reciprocal geodesic distances [49]. Thus, the higher values indicate more central position. The normalized closeness centrality, $C_C^{'}(n_i)$ which ranges from 0 to 1, is found by multiplying $C_C(n_i)$ by *n*-1. Then, $C_C^{'}(n_i)$ is given by:

$$C_C^{'}(n_i) = \frac{n-1}{\sum_{j}^{n} \mathrm{d}(n_i, n_j)}. \tag{2.13}$$

Betweenness Centrality of a node n_i, denoted by $C_B(n_i)$, is the sum of the ratio of the number of geodesics, $g_{jk}(n_i)$, linking the nodes n_j and n_k that contain node n_i to the number of geodesics, g_{jk}, linking the nodes n_j and n_k [49]. In other words, it counts "the number of geodesic paths (i.e., shortest paths) that pass through a node n_i [116]. The normalized betweenness centrality, $C_B^{'}(n_i)$ which ranges from 0 to 1, is found by dividing the betweenness centrality by (n-1)(n-2)/2 which indicates the number of pairs of nodes not including n_i. Then, $C_B^{'}(n_i)$ is given by:

$$C_B^{'}(n_i) = \frac{C_B(n_i)}{(n-1)(n-2)/2} = \sum_{j<}^{n} \sum_{k}^{n} \frac{g_{jk}(n_i)}{g_{jk}}. \tag{2.14}$$

Eigenvector Centrality a node n_i, denoted by $C_E(n_i)$, is a variant of degree centrality in which a node is more central if it is connected to nodes that are themselves well-connected [51, 117]. It is computed by solving:

$$A * c = \lambda * c, \tag{2.15}$$

where *A* is the adjacency matrix for a graph in which $a_{ij} = 1$ if vertex *i* is connected to vertex *j*, and $a_{ij} = 0$ otherwise, *c* is a vector of the degree centralities for each vertex as indicated by

$c = (C_D(n_1), C_D(n_2),..., C_D(n_n))'$, and λ is a scalar. The above equation is the characteristic equation to find the eigensystem of a matrix A [49]. Then, the elements of eigenvector are the eigenvector centralities, $C_E(n_i)$, for each vertex of the graph. By convention, eigenvector centrality is given by the eigenvector with the largest eigenvalue λ [89]. The normalized eigenvector centrality, $C_E^{'}(n_i)$ can be found by "the square root of one half, which is the maximum score attainable in any graph" [51, 118]. Then, $C_E^{'}(n_i)$ is given by:

$$C_E^{'}(n_i) = C_E(n_i) \big/ \sqrt{2}\ . \tag{2.16}$$

Table 2.14 illustrates the descriptive statistics for all type of centrality metrics across four networks. The network of communication had the highest mean value for all type of centrality metrics, except for betweenness centrality which had the second lowest mean value. This indicated that there were not, on average, lots of researchers who played a brokerage or gatekeeper role in the network of communication. The network of joint grant proposals had higher mean value for all type of centrality metrics than the network of joint publications, except for eigenvector centrality that was lower in the network of joint grant proposals. This implied that the researchers' tendencies to publish results with other researchers that were well-connected were, on average, more than their tendencies to write grant and submit proposals with other researchers that were well-connected.

It is also important to analyze the degree to which a whole network has a centralized structure. Table 2.15 illustrates the network centralization which measures the degree of inequality or variance in a network as a percentage of a perfect star network of the same size [49, 93, 119]. In other words, the graph centralization measures how tightly a network is organized around its most central node [120]. In the network of communication, there was a significant amount of degree centralization in the whole network when compared to the collaborative output networks. This implied that the degree centrality of individual nodes significantly varied and the advantages arising from degree centralities were distributed unequally in the network of communication [93]. The value of closeness centralization for the network of communication and joint publications were very close to each other and higher than the other networks. Overall, the values for closeness centralization indicated that closeness centrality of the individual nodes varied in all network, especially in the network of communication and joint publications. Betweenness centralization for all networks was low, indicating that the values for betweenness centrality of the individual nodes were evenly distributed in all networks. The network of joint publications had the highest value for eigenvector centralization, meaning that eigenvector centrality of the individual nodes varied in the network of joint publications compared to other networks.

The degree centrality of researchers who share the same attributes was also analyzed. Table 2.16 illustrates the normalized group degree centralities. While calculating the group centralities, the groups such as such as gender, race, and department affiliation are treated as one node, and its ties to other nodes are computed. The multiple ties from other nodes to this node are counted only once [121]. Males were more central than Females in all networks. The centrality of different races was ranged from high to low in all networks as follows: White, Asian, Hispanic, and Black, except that Blacks were more central than Hispanics in the network of joint patents. In the network of communication, the most central department was CBE, whereas the least one was CEE. In the network of joint publications, the most and least central departments were EE and IMSE, respectively. In the network of joint grant proposals, the highest centrality was scored by EE, while the lowest centrality was scored by IMSE. In the network of joint patents, CBE had the highest group centrality, whereas CSE had the lowest group centrality.

The difference in the means of group centralities was tested as well. Table 2.17a illustrates the results for the comparison of the means of group centralities in each network. While a t-test was run for comparing two groups in 'gender' attribute, one-way analysis of variance (ANOVA) was run for comparing multiple groups in 'race' and 'department affiliation' attribute. For both methods, since the observations were not independents a method called random sampling of permutations were used to calculate an approximate p-value. To create the permutation based sampling distribution of the difference between both the means of two groups and multiple groups, large number of trials were run (in this study, 10000 for two groups and 5000 for multiple groups) [93]. In each trial, centrality scores for each individual were randomly assigned to another individual; that is, they were randomly permuted. Standard deviation of the distribution created by random trials became estimated standard error for t-test and ANOVA [93]. If the difference in the means of group centralities was statistically significant it was bolded in red in Table 2.17a. Moreover, R-square values ranged from 0.001 to 0.210 were given in Table 2.17b for multiple group comparisons. For two groups, the only statistically significant difference was in the mean of both male and female eigenvector centralities in the network of joint grant proposals. This implied that the connections of males and females to other well-connected researchers were different in the network of joint grant proposals. For multiple groups, there was significant difference in the means of betweenness and degree centralities of races in the network of joint publications. This implied that both the number of researchers' direct connections to other researchers and the number of researchers who locate themselves in shortest paths showed difference among the races in the network of joint publications. Moreover, there were significant differences in the mean of eigenvector centralities of department affiliations in all networks. This implied that the researchers in some departments tended to be connected to other well-connected researchers much more in all networks.

2.5. Discussion

This study demonstrates how comparative analysis of researchers' communication and collaborative output networks (e.g., network of joint publications, grant proposals, and patents) is performed in the presence of self-reported data collected in a college of engineering. It presents a data collection method that enables us not only to collect the frequency of communication between researchers but also to collect the self-report of the number of in-progress and completed collaborative outputs between researchers.

The method facilitates the comparative analysis of researchers' communication and collaborative output networks by using a richer dataset taking into account both in-progress and collaborative efforts. Collecting researchers' collaborative output data in a self-reported way provides some indication of whether or not a tie is important in terms of their collaborative research efforts. In other words, the self-reported way of collecting the relations in collaborative outputs permits the researchers to assess both which connection or tie is important to them according to their own perceptions and whether or not reported contact is actually involved in research. Furthermore, collecting relational data simultaneously for multiple networks helps us to understand the extent to which the structure of these networks overlaps and the extent to which researchers' communication relations impact their collaboration relations from the network perspective. That is, gathering data for researchers' informal conversational exchange ties and collaborative output ties with other researchers simultaneously helps to test not only the extent to which researchers' collaborative output ties can be really used as a proxy for their communication ties but also the extent to which scientific collaboration is nurtured by means of informal conversational exchange [18, 33].

Table 2.1. Advantages of Scientific Collaboration

Access to expertise for complex problems, new resources and, funding	[6-13]
Increase in the participants' visibility and recognition	[8, 10]
Rapid solutions for more encompassing problems by creating a synergetic effect among participants	[10, 14]
Decrease in the risks and possible errors made, thereby increasing accuracy of research and quality of results due to multiple viewpoints	[10, 11]
Growth in advancement of scientific disciplines and cross-fertilization across scientific disciplines	[10, 15]
Development of the scientific knowledge and technical human capital, e.g., participants' formal education and training, and their social relations and network ties with other scientists	[16]
Increase in the scientific productivity of individuals and their career growth	[8, 16-18]

Table 2.2. Number of Researchers in Each Demographic Attribute

	Gender						Total
	Male			Female			
sample	86			14			100
participants	68			8			76
	Race						
	Asian		Black	Hispanic	White		
sample	35		4	9	52		100
participants	28		3	5	40		76
	Department						
	CBE	CEE	CSE	EE	IMSE	ME	
sample	16	19	17	24	10	14	100
participants	14	13	10	17	10	12	76

Table 2.3. Timeline of the Steps Performed During the Data Collection

Timeline	Steps
During the first week of October, 2012	A pilot test conducted for the questionnaire.
In the middle of October, 2012	A mass e-mail from the dean's office was sent out to inform the researchers.
During the last two weeks of October, 2012	Questionnaires began to be distributed either in the departmental meetings or through in-person delivery and e-mail.
During the first week of November, 2012	A follow-up e-mail was sent to collect the completed questionnaires. The response rate was very low. Therefore, questionnaires were delivered to the researchers in person intensively. An extra one week was given to the participants for uncompleted questionnaires
During the second week of November, 2012	Completed questionnaires continued to be collected, and also the questionnaires continued to be delivered in person.
During the last week of November and December, 2012	Due to the holiday season, there was minimum response received from the researchers.
In the first week of March, 2013	All responses from the participants were finalized.

Table 2.4. Five Possible Cases of Reciprocity

Cases	**Upper Triangle Cells**	**Lower Triangle Cells**
1	Equal	Equal
2a	High	Low
2b	Low	High
3a	X	0
3b	0	X

Table 2.5. The Number of Occurrences of Five Possible Cases in Each Network and Inter-rater Agreement Percentage

Cases	**Network of Communication**	**Network of Joint Publications**	**Network of Joint Grant Proposals**	**Network of Joint Patents**
1	120	38	81	9
2a	141	14	20	2
2b	144	16	21	2
3a	377	68	113	11
3b	452	60	132	11
Inter-rater agreement percentage	9.72%	19.39%	22.07%	25.71%

'1' The value of the upper and the lower triangle cells were equal.
'2a' The value of the upper triangle cells was higher than the value of the lower triangle cells.
'2b' The value of the lower triangle cells was higher than the value of the upper triangle cells.
'3a' The upper triangle cells contained a value, but lower triangle cells did not.
'3b' The lower triangle cells contained a value, but the upper triangle cells did not.

Table 2.6. The Most Idealistic Scenario of the Conversion to Undirected Edges

Cases	**Upper Triangle Cells**	**Lower Triangle Cells**
1	Equal	Equal
2a*	High	High
2b*	High	High
3a*	X	X
3b*	X	X

Table 2.7. Statistical and Descriptive Properties of Four Networks

	Network of Communication	Network of Joint Publications	Network of Joint Grant Proposals	Network of Joint Patents
Vertices (active)	100	90	97	35
Total Edges	1234	196	367	35
Connected Components (or CCs)	1	1	1	7
Single-Vertex CCs	0	10	3	65
Maximum Vertices in a CC	100	90	97	23
Maximum Edges in a CC	1234	196	367	29
Graph Density (Binary)	0.249	0.040	0.074	0.007
Graph Density (Valued)	0.741	0.068	0.107	0.010
Maximum Geodesic Distance (or Diameter) in a CC	3	7	7	9
Average Geodesic Distance	1.792	3.468	2.699	3.452
Clustering coefficient	0.534	0.158	0.285	0.051
Assortativity(Degree)	-0.146	0.044	0.072	0.217
Distance-based cohesion ("Compactness")	0.618	0.277	0.401	0.021
Distance-weighted fragmentation ("Breadth")	0.382	0.723	0.599	0.979
Number of collaborators per researcher	12.34	1.96	3.67	0.35

Note: This table was constructed by means of three computer packages: NodeXL version 1.01.229, UCINET 6.308, and The R project for statistical computing.

Table 2.8. Comparison of Network Densities

		St. Dev. Diff. by Classical Method	Mean Diff. by Classical Method	St. Dev. Diff. by Bootstrap Sampling	Mean Diff. by Bootstrap Sampling
			Binary relations		
Communication	Joint Publications	0.005	0.209	0.014	0.207*
	Joint Grant Proposals	0.005	0.175	0.012	0.176*
	Joint Patents	0.004	0.242	0.015	0.239*
Joint Publications	Joint Grant Proposals	0.003	0.034	0.006	0.034*
	Joint Patents	0.002	0.033	0.004	0.032*
Joint Grant Proposals	Joint Patents	0.003	0.067	0.007	0.066*
			Valued relations		
Joint Publications	Joint Grant Proposals	0.006	0.039	0.010	0.039*
	Joint Patents	0.004	0.058	0.007	0.057*
Joint Grant Proposals	Joint Patents	0.005	0.097	0.011	0.096*

*<0.01 Note: 10000 Bootstrap samples

Table 2.9. QAP Correlation between Networks

	Communication	Joint Publications	Joint Grant Proposals	Joint Patents
Jaccard coefficient for binary relations				
	Communication	**Joint Publications**	**Joint Grant Proposals**	**Joint Patents**
Communication	1.000	0.154*	0.283*	0.028*
Joint Publications		1.000	0.328*	0.155*
Joint Grant Proposals			1.000	0.072*
Joint Patents				1.000
Pearson's correlation for valued relations				
	Communication	**Joint Publications**	**Joint Grant Proposals**	**Joint Patents**
Communication	1.000	0.366*	0.484*	0.154*
Joint Publications		1.000	0.600*	0.447*
Joint Grant Proposals			1.000	0.317*
Joint Patents				1.000

*<0.01, Note: 5000 permutations were run for QAP.

Table 2.10. QAP Correlation (Pearson's Correlation for Valued Relations) between Researchers' Spatial Proximity and Their Multiple Networks

Networks	Spatial Proximity
Communication	0.384*
Joint Publications	0.118*
Joint Grant Proposals	0.140*
Joint Patents	0.047*

*<0.01, Note: 5000 permutations were run for QAP.

Table 2.11. QAP Regression of Researchers' Communication on Their Collaborative Output Networks (Binary Relations)

Networks	Standardized beta coefficients	QAP significance	R-square	p-value
Joint Publications (dependent network)				
Communication	0.128	<.001	0.324	<.001
Joint Grant Proposals	0.373	<.001		
Joint Patents	0.264	<.001		
Joint Grant Proposals (dependent network)				
Communication	0.337	<.001	0.345	<.001
Joint Publications	0.362	0.002		
Joint Patents	0.043	<.001		
Joint Patents (dependent network)				
Communication	0.006	0.345	0.137	<.001
Joint Publications	0.337	<.001		
Joint Grant Proposals	0.056	0.003		

Note: 10000 permutations were run for QAP.

Table 2.12. QAP Regression of Researchers' Communication on Their Collaborative Output Networks (Valued Relations)

Networks	Standardized beta coefficients	QAP significance	R-square	p-value
Joint Publications (dependent network)				
Communication	0.099	<.001	0.440	<.001
Joint Grant Proposals	0.462	<.001		
Joint Patents	0.285	<.001		
Joint Grant Proposals (dependent network)				
Communication	0.305	<.001	0.444	<.001
Joint Publications	0.459	<.001		
Joint Patents	0.065	<.001		
Joint Patents (dependent network)				
Communication	-0.040	<.001	0.204	<.001
Joint Publications	0.405	<.001		
Joint Grant Proposals	0.093	<.001		

Note: 10000 permutations were run for QAP.

Table 2.13. Exponential Random Graph Models (ERGMs) to Predict the Properties of Networks

Joint Publications (as dependent network)										*****<0.001, **<0.01, *< 0.05**
	Model 1		**Model 2**		**Model 3**		**Model 4**		**Model 5**	
	Estimates	Std.	Estimates	Std.	Estimates	Std.	Estimates	Std.	Estimates	Std.
Edges	-3.188***	0.052	-4.678***	0.106	-4.487***	0.150	-4.936***	0.167	-5.376***	0.196
Communication			0.756***	0.027			0.774***	0.034	0.474***	0.044
Gender (Common)					0.381**	0.138	0.145	0.147	0.227	0.170
Race(Common)					0.477***	0.107	0.387***	0.115	0.519***	0.136
Department (Common)					1.719***	0.120	-0.035	0.144	0.231	0.169
Next to each other					0.268	0.355	-1.064**	0.376	-0.920*	0.453
The same hallway					0.758***	0.135	-0.024	0.145	0.320	0.173
The same building					0.112	0.149	-0.167	0.160	-0.111	0.190
Different buildings					NA	0.000	NA	0.000	NA	0.000
Joint Grant Proposals									1.549***	0.092
Joint Patents									3.397***	0.404
AIC	3302		2360		2911		2348		1827	
Joint Grant Proposals (as dependent network)										
	Model 1		**Model 2**		**Model 3**		**Model 4**		**Model 5**	
	Estimates	Std.	Estimates	Std.	Estimates	Std.	Estimates	Std.	Estimates	Std.
Edges	-2.525***	0.038	-3.753***	0.069	-3.212***	0.097	-3.586***	0.109	-3.573***	0.113
Communication			0.730***	0.020			0.840***	0.028	0.753***	0.031
Gender (Common)					-0.001	0.093	-0.276**	0.104	-0.381***	0.110
Race (Common)					0.352***	0.080	0.255**	0.090	0.177	0.097
Department (Common)					1.324***	0.091	-0.590***	0.122	-0.644***	0.132
Next to each other					0.808**	0.253	-0.640*	0.293	-0.540	0.315
The same hallway					0.695***	0.105	-0.187	0.121	-0.194	0.132
The same building					0.071	0.108	-0.243*	0.122	-0.277*	0.133
Different buildings					NA	0.000	NA	0.000	NA	0.000
Joint Publications									1.670***	0.112
Joint Patents									-0.136	0.339
AIC	5234		3777		4809		3740		3349	
Joint Patents (as dependent network)										
	Model 1		**Model 2**		**Model 3**		**Model 4**		**Model 5**	
	Estimates	Std.	Estimates	Std.	Estimates	Std.	Estimates	Std.	Estimates	Std.
Edges	-4.945***	0.120	-6.463***	0.253	-6.610***	0.382	-7.096***	0.415	-7.142***	0.458
Communication			0.713***	0.059			0.733***	0.074	0.224*	0.102
Gender (Common)					0.623	0.346	0.415	0.350	0.032	0.384
Race (Common)					0.404	0.243	0.282	0.246	0.425	0.297
Department (Common)					1.619***	0.273	-0.126	0.301	0.253	0.367
Next to each other					1.326*	0.578	0.179	0.588	1.042	0.700
The same hallway					0.662*	0.336	-0.094	0.334	0.306	0.399
The same building					0.911**	0.301	0.701*	0.306	1.112**	0.355
Different buildings					NA	0.000	NA	0.000	NA	0.000
Joint Grant Proposals									0.569**	0.175
Joint Publications									1.123***	0.135
AIC	834		667		770		669		485	

Table 2.14. Mean and Standard Deviation of Four Centrality Types (Normalized)

Networks	Degree		Closeness		Betweenness		Eigenvector	
	Mean	**St. Dev.**	**Mean**	**St. Dev.**	**Mean**	**St. Dev.**	**Mean**	**St. Dev.**
Communication	0.249	0.116	0.618	0.065	0.008	0.011	0.130	0.057
Joint Publications	0.040	0.031	0.277	0.106	0.020	0.035	0.094	0.106
Joint Grant Proposals	0.074	0.056	0.401	0.099	0.097	0.103	0.016	0.022
Joint Patents	0.007	0.013	0.021	0.037	0.001	0.004	0.048	0.133

Table 2.15. Network Centralization

Networks	Degree	Closeness	Betweenness	Eigenvector
Communication	40.53%	41.70%	6.79%	19.91%
Joint Publications	13.48%	37.37%	17.66%	56.68%
Joint Grant Proposals	16.14%	31.64%	9.21%	10.85%
Joint Patents	7.52%	23.66%	3.05%	5.43%

Table 2.16. Normalized Group Degree Centralities

	Gender					
Networks	**Male**			**Female**		
Communication	1.000			0.953		
Joint Publications	0.857			0.279		
Joint Grant Proposals	0.929			0.500		
Joint Patents	0.214			0.058		
	Race					
	Asian	**Black**	**Hispanic**	**White**		
Communication	1.000	0.646	0.868	1.000		
Joint Publications	0.646	0.125	0.275	0.688		
Joint Grant Proposals	0.723	0.271	0.352	0.833		
Joint Patents	0.138	0.052	0.033	0.167		
	Department					
	CBE	**CEE**	**CSE**	**EE**	**IMSE**	**ME**
Communication	0.857	0.691	0.747	0.789	0.833	0.744
Joint Publications	0.250	0.185	0.229	0.329	0.167	0.314
Joint Grant Proposals	0.476	0.346	0.313	0.566	0.300	0.547
Joint Patents	0.095	0.012	0.000	0.079	0.011	0.081

Table 2.17a. Hypothesis Test about Mean Centrality of Groups in Each Network

Centrality	Two groups[1] (Gender)			Multiple groups[2] (Race)					Multiple groups[2] (Department)						
Communication															
	Male	**Female**	**p-value[3]**	**Asian**	**Black**	**Hispanic**	**White**	**F-value**	**CBE**	**CEE**	**CSE**	**EE**	**IMSE**	**ME**	**F-value**
Betweenness	0.009	0.008	0.713	0.008	0.007	0.008	0.008	0.021	0.012	0.006	0.007	0.007	0.010	0.008	0.860
Closeness	0.622	0.617	0.806	0.621	0.625	0.638	0.611	0.473	0.645	0.593	0.600	0.620	0.638	0.623	1.645
Degree	0.255	0.248	0.857	0.254	0.250	0.280	0.241	0.308	0.299	0.210	0.218	0.253	0.281	0.258	1.456
Eigenvector	0.130	0.130	0.983	0.131	0.132	0.151	0.125	0.555	0.159	0.103	0.111	0.133	0.147	0.138	**2.548***
Joint Publications															
	Male	**Female**	**p-value**	**Asian**	**Black**	**Hispanic**	**White**	**F-value**	**CBE**	**CEE**	**CSE**	**EE**	**IMSE**	**ME**	**F-value**
Betweenness	0.018	0.021	0.827	0.033	0.013	0.040	0.009	**5.090***	0.013	0.012	0.023	0.020	0.045	0.021	1.390
Closeness	0.254	0.281	0.390	0.311	0.236	0.286	0.256	2.100	0.238	0.268	0.284	0.302	0.308	0.263	0.952
Degree	0.035	0.040	0.595	0.050	0.035	0.043	0.032	**2.522****	0.041	0.034	0.038	0.044	0.041	0.040	0.250
Eigenvector	0.081	0.096	0.621	0.121	0.101	0.082	0.077	1.229	0.142	0.052	0.063	0.126	0.061	0.101	**2.312***
Joint Grant Proposals															
	Male	**Female**	**p-value**	**Asian**	**Black**	**Hispanic**	**White**	**F-value**	**CBE**	**CEE**	**CSE**	**EE**	**IMSE**	**ME**	**F-value**
Betweenness	0.017	0.016	0.838	0.021	0.012	0.012	0.014	0.696	0.022	0.011	0.013	0.017	0.072	0.017	0.559
Closeness	0.409	0.400	0.752	0.414	0.433	0.415	0.388	0.708	0.393	0.406	0.363	0.415	0.399	0.429	0.846
Degree	0.091	0.071	0.236	0.081	0.091	0.065	0.070	0.451	0.095	0.071	0.057	0.075	0.062	0.083	0.952
Eigenvector	0.159	0.086	**0.015***	0.100	0.164	0.084	0.092	0.645	0.168	0.117	0.031	0.092	0.049	0.107	**3.990***
Joint Patents															
	Male	**Female**	**p-value**	**Asian**	**Black**	**Hispanic**	**White**	**F-value**	**CBE**	**CEE**	**CSE**	**EE**	**IMSE**	**ME**	**F-value**
Betweenness	0.002	0.001	0.480	0.001	0.005	0.001	0.001	0.822	0.002	0.000	0.000	0.004	0.000	0.001	**2.207***
Closeness	0.017	0.022	0.687	0.018	0.040	0.010	0.024	0.795	0.038	0.004	0.004	0.043	0.008	0.017	**5.011***
Degree	0.005	0.007	0.611	0.005	0.013	0.003	0.008	0.772	0.011	0.002	0.004	0.014	0.001	0.007	**3.027***
Eigenvector	0.015	0.054	0.332	0.041	0.042	0.004	0.061	0.517	0.056	0.002	0.000	0.128	0.013	0.050	**3.074***

*<0.05, **<0.10 Note: significant values are in red . [1]Hypotheses were tested by t-test (permutation by 10000 trials). [2]Hypotheses were tested by ANOVA (permutation by 5000 trials). [3]UCINET version 6.308 does not provide t-test statistics results.

Table 2.17b. R-square Values of ANOVAs for Multiple Groups

	Race				Department			
	Communication	**Joint Publications**	**Joint Grant Proposals**	**Joint Patents**	**Communication**	**Joint Publications**	**Joint Grant Proposals**	**Joint Patents**
Betweenness	0.001	0.137	0.021	0.025	0.044	0.069	0.029	0.105
Closeness	0.015	0.062	0.022	0.024	0.080	0.048	0.043	0.210
Degree	0.010	0.073	0.014	0.024	0.072	0.013	0.048	0.139
Eigenvector	0.017	0.037	0.020	0.016	0.119	0.110	0.175	0.141

CHAPTER 3: A REGRESSION ANALYSIS OF RESEARCHERS' SOCIAL NETWORK METRICS ON THEIR CITATION PERFORMANCE IN A COLLEGE OF ENGINEERING

3.1. Introduction

It is important to determine who are the most influential researchers and invest in those researchers to both maximize the research outputs and to allocate funding effectively [51, 122]. Influential researchers can be determined by using social network metrics such as centrality metrics after mapping their collaborative output networks (e.g., joint publications, grant proposals, and patents) in which a tie between any two authors indicates collaboration on the making of a collaborative output. Hou et al. (2008) found that there was a positive correlation between being an influential researcher, (i.e., having a high degree centrality in the collaborative output network) and output of a researcher (i.e., number of publications). Defazio et al. (2009) also found that there was high impact of being an influential researcher in the collaborative output network on output of a researcher. However, the quality of research outputs is as important as the quantity of the research outputs.

Hirsch (2005) proposed an index called the h-index in order to attempt to measure both the number of publications a researcher produced (i.e., quantity) and their impact on other publications (i.e., quality). Using the researchers' publications data in the information schools of five universities, Abbasi et al. (2011) investigated the impact of social network metrics (including different centrality metrics, average tie strength, and efficiency coefficient proposed by Burt (1992)) obtained from a researchers' co-authorship network on the their g-index (another form of h-index), and found out that degree centrality, average tie strength, and efficiency coefficient had a positive impact on the researchers' performance, while eigenvector centrality had a negative impact on the researchers' performance. Their study can be extended by considering the network metrics obtained from researchers' multiple networks. Thus, the purpose of this study is to test the findings of Abbasi et al. (2011) with the social network metrics obtained from researchers' multiple collaborative networks defined by joint publications, joint grant proposals, and joint patents as well as their communication network to understand the relationship between these social network metrics and the performance of researchers. Collecting researchers' ties for their informal conversational exchange (or informal communication) and collaborative outputs with other researchers within a college simultaneously makes this testing possible. This study uses h-index instead of the g-index because the researchers within the same field of study are compared [124]. In sum, this study seeks an answer to the following question: *what is the impact of social network metrics obtained from researchers' communication and collaborative output networks on their performance as measured by citations of their publications?*

3.2. Literature Review and Hypotheses

3.2.1. A Performance Measure of Researchers: h-index

A researcher's performance is assessed by two factors: the number of publications he/she produced and the impact of those publications in the scientific community [124-126]. Hirsch (2005) proposed an index called h-index that combined both of these quantity and impact factors. The h-index drew the attention of many researchers in the scientific community, and many publications on this topic emerged [126]. Hirsch (2005) defined the h-index as follows: "A scientist has index h if h of his/her N_p papers have at least h citations each and the other (N_p-h) papers have fewer than h citations each, where N_p is the number of papers published over n years" [127]. Even though the h-index was better than straight citation counts [127] and had more predictive power to assess the future achievement of researchers [128], different modifications of the h-index have been proposed in the literature to overcome its shortcomings [126, 129]. Some shortcomings are as follows: favoring disciplines which do experimental research study in larger groups such as physics, assigning an equal value to each author in multiple-author papers, not accounting for author sequence and the total number of authors, being inflated via self-citations, not considering books and other alternative forms of publication, not considering the performance changes throughout a researcher's career and lag time between a paper being published and being discovered and cited [130]. In this study, the h-index, the most widely used performance metric for researchers, was used because the researchers within the same field of study are compared [124].

3.2.2. Social Network Metrics

Sonnenwald (2007) defined scientific collaboration as the interaction within a social context among two or more scientists in order to facilitate the completion of tasks with regard to a commonly shared goal. Thus, those collaborations are perpetuated through social networks [51]. SNA is the method used to reveal the structure of collaboration between individuals [42, 43]. Hence, many social network metrics in SNA are used to analyze the structure of collaboration between researchers [25-27, 79]. Using the data gathered by the questionnaire, the goal of this study is to test the impact of the following social network metrics extracted from both researchers' communication and collaborative output networks on the researchers' citation-based performance index (h-index).

- *Degree Centrality* (i.e., the researchers' distinct connections to many different researchers)
- *Closeness Centrality* (i.e., the shortness of a researcher's total distance to all other researchers)
- *Betweenness Centrality* (i.e., the number of times the researchers holding the shortest path between two other researchers)
- *Eigenvector Centrality* (i.e., the researcher's tendency to connect with other researchers who are themselves well-connected)
- *Average Tie Strength* (i.e., the researcher's averaged number of repeated collaborative outputs with other researchers)
- *Burt's Efficiency Coefficient* (i.e., the researchers' redundant connections to a group of researchers who are themselves well-connected)
- *Local Clustering Coefficient* (i.e., an researcher's tendency towards the dense local neighborhoods)

The discussion for *degree centrality*, *closeness centrality*, *betweenness centrality*, *eigenvector centrality*, and *local clustering coefficient* was already made in section 2.4.5. Therefore, this chapter only discusses the following two social network metrics: *average tie strength* and *efficiency coefficient*. Unlike the study of Abbasi et al. (2011), this study also considers the local clustering coefficient which is an individual's tendency towards the dense local neighborhoods. The local clustering coefficient is also defined as a measure of degree to which an individual is embedded in a tightly knit groups, i.e., positioned in a dense-connected cluster [93, 131]. It is necessary to consider the local clustering coefficient of a researcher because it is more likely that working in a team (or being in dense-connected cluster) leads to higher number of citations [78, 132]. Therefore, the impact of the researchers' tendency towards the dense local neighborhoods on their citation performance (h-index) is tested.

Average Tie Strength of a node n_i, denoted by *ATS*, is the proportion of the sum of unique weighted edges (the strength of a tie or an edge as the weight of the edge) that are connected to node n_i to the number of unique edges connected to node n_i (i.e., degree centrality of the node, $C_D(n_i)$). Then, similar to the calculation in Abbasi et al. (2011), for the network of collaborative outputs, *ATS* is calculated; by dividing a researcher's total number of collaborative outputs, *NCO*, with other researchers by the number of his/her reported collaborators. For the network of communication, it is calculated by dividing a researcher's total conversational exchange frequencies with other researchers, *TF*, by the number of his/her reported conversational partners. Then, the average tie strength is given by:

$$ATS(n_i) = \frac{\sum_k^n NCO_{ik}}{C_D(n_i)} \text{ or } \frac{\sum_k^n TF_{ik}}{C_D(n_i)} \tag{3.1}$$

Efficiency coefficient proposed by Burt (1992) considers the redundancy of an individual's contacts [133]. The theory of structural holes claims that the case that an individual (or ego) is connected to an individual who is in a close-knit group is more advantageous than the case that an individual is connected to several individuals who are in the same close-knit group [52, 133]. The main reason for this is that the connections to several individuals in the close-knit group creates redundancy to the ego since information benefits provided by an individual in the close-knit group are redundant with benefits provided by other individual in the close-knit group [52]. Burt's efficiency coefficient for non-valued and undirected relations is given by:

$$Ef(n_i) = \frac{\sum_j m_{ij} \left[1 - \sum_q p_{iq} m_{jq} \right]}{\sum_j z_{ij}}, \tag{3.2a}$$

where p_{iq} is the proportion of node *i*'s network time and energy invested in the relationship with node *q* (node *i*'s contact) and calculated by:

$$p_{iq} = \frac{z_{iq}}{\sum_j z_{ij}}, \quad i \neq j, \tag{3.2b}$$

where z_{iq} is the strength of the relationship between node *i* and *q* (in binary case, 1), and $\sum_j z_{ij}$ is the total strength of the relationship with *j* contacts [52, 133]. m_{jq} is the marginal strength of contact *j*'s relation with contact *q* and calculated by:

$$m_{jq} = \frac{z_{jq}}{\max_k z_{jk}}, \quad j \neq k, \tag{3.2c}$$

where $\max_k z_{jk}$ is the largest of *j*'s relations with anyone, and z_{jq} is the strength of the relations from *j* to *q* [52, 133]. Since $\max_k z_{jk}$ is 1 in non-valued and undirected graph, it becomes m_{jq} = z_{jq} [52, 133].

The impact of social network metrics on the performance of individuals can be found in many studies using different types of communication and collaborative networks, e.g., the positive impact of closeness centrality in the communication network of M.B.A. students on their grade performances [134], the positive impact of betweenness centrality in both friendship network and workflow network of employees in a small high-technology company on their workplace performance [135], the positive impact of degree centrality and network density in the advice network of employees in 5 different organizations on individual job performance and group performance [136], and the positive impact of eigenvector centrality of group leaders in their friendship networks in the sales division of a financial services firm on the performance of their groups [137]. Then, based on the definition of social network metrics discussed so far, the following 7 hypotheses about the impact of a researcher's position on his/her performance are tested for each network, namely the communication network, the network of joint publications, grant proposals, and patents.

Hypotheses 1 to *7*: The network metrics in terms of researchers' degree centrality (1), closeness centrality (2), betweenness centrality (3), eigenvector centrality (4), average tie strength (5), efficiency measure (6), and local clustering coefficient (7) positively impact their citation performance (e.g., h-index).

3.3. Method

3.3.1. Constructing Data Sets for Statistical Model

Four datasets from four social network data matrixes corresponding to researchers' each network (e.g., communication, joint publications, joint grant proposals, and joint patents) were constructed. Each of four datasets included 11 variables for 100 researchers. In other words, four data matrixes in 100x11 dimensions were compiled. The variables included in the datasets are the researchers' citation-based performance index (h-index), 7 social network metrics obtained from each network (i.e., degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, average tie strength, Burt's efficiency coefficient, and local clustering coefficient), and 3 demographic attributes (i.e., gender, race, and department affiliation). The researchers' citation-based performance index (h-index) can be easily obtained through the Thomson ISI Web of Science database without the need for further calculation [125]. The database was accessed via the library of the University of South Florida. Each researcher's h-index was obtained by plugging the researcher's name, an organization name (e.g., the University of South Florida), and the years between 2006 and 2012 into the search boxes. The social network metrics for each network were computed using UCINET 6.308 [89]. While centrality metrics, Burt's efficiency coefficient, and local clustering coefficient were computed using dichotomized data matrixes, average tie strength were computed using valued data matrixes.

3.3.2. Poisson Regression Model

Poisson regression is one of the standard (or base) count response regression models [138]. It can be used in many different fields such as health services (e.g., doctor visits), finance and economics (e.g. recreational demands, takeover biddings, bank failures, accidental insurances, and credit ratings), political science (e.g., presidential appointments), informetrics (patents, doctoral publications) and so forth [139]. The Poisson regression models were run in this study because h-index is count data, and the mean and variance of the variable h-index was reasonably close to each other (Mean=3.47 and Variance=2.78). The general form of a Poisson regression model is given by:

$$E(Y = y \mid \mathrm{x}) \; or \; \mu = \exp(x'\beta) = \exp(x_1\beta_1)\exp(x_2\beta_2)\ldots\exp(x_n\beta_n)$$
$$\text{or}$$
$$\log(\mu) = x'\beta \tag{3.3}$$

where y is the dependent variable, μ is the mean, and x' and β are the linearly independent repressors and regression coefficients. In the abovementioned model, the multiplicative effect of predictor, x_j on the mean is represented by the exponentiated regression coefficient, $\exp(\beta_j)$. One unit increase in x_j multiplies the mean by a factor of $\exp(\beta_j)$ [140]. The main reason for log transformation is to keep the left hand side of the equation that indicates an expected count non-negative [139]. The multicollinearity problem occurs when there is a high correlation among two or more of the independent variables in a multiple regression, meaning that one independent variable or predictor can be predicted from others [141]. This problem can be even more explicit when social network metrics are used as predictors. The Spearman's rank correlations in Table 3.1 indicate that many of social network metrics, especially centrality metrics, are extremely correlated. Running a multiple regression with these highly correlated social network metrics as predictors gives unreliable estimates about an individual predictor. To overcome the challenge of potential multicollinearity between predictors, this study run a separate Poisson regression bivariate model for each of seven SNA metric obtained from each network. Then, the models that were run for different SNA metrics in each network can be shown by:

$$\log(h-index) = \beta_0 + \beta_1(a\ SNA\ metric) + \beta_2 Gender + \beta_3 Race + \beta_4 Department \tag{3.4}$$

Analysis for the models was performed using IBM SPSS Statistics for Windows, Version 21.0. (Armonk, NY: IBM Corporation).

3.4. Results

Table 3.2 illustrates the bivariate model results for each network. Maximum likelihood estimation was used to estimate the regression coefficients of predictors (or parameters) in the model. Likelihood Ratio Chi-Square test (also called omnibus test or test against the intercept-only model) evaluates whether or not all of the estimated coefficients are equal to zero; in other words, it is the test of the model as whole [142]. From the p-values, all models were statistically significant at the significance level of 0.05.

The estimated regression coefficients for each network parameter indicated the following results. Degree centrality (C_D) was statistically significant and had a positive impact in all networks except the communication network. Unlike the results of Abbasi et al. (2011), closeness centrality (C_C) and eigenvector centrality (C_E) were statistically significant, and had a positive impact on the citation performance for all networks. Betweenness centrality (C_B) had a positive significant impact for only the network of joint publications. Average tie strength (*ATS*) was statistically significant, and had a positive impact for only the network of joint publications and patents. Efficiency coefficient (*Ef*) had a positive significant impact for only the network of patents. The local clustering coefficient (*LCC*) was statistically significant and had a positive impact for only the network of joint publications and grant proposals.

The Poisson regression coefficients are interpreted as follows: "for a one unit change in the predictor variable, the difference in the logs of expected counts is expected to change by the respective regression coefficient, given the other predictor variables in the model are held constant." [142]. For example, if a researcher in the College of Engineering increases his/her eigenvector centrality score (i.e., increase his her/her connections with the researchers who are well connected) by one point in the network of communication, joint publications, joint grant proposals, and joint patents, the difference in the logs of expected h-index is expected to increase by a factor of 3.345, 3.212, 2.956, and 1.306, respectively, while the other variables are held constant in the model. The coefficients can also be exponentiated to assess the relationship between the response and predictors as incidence rate ratios (IRR) [138]. For one unit increase in eigenvector centrality scores in the network of communication, joint publications, joint grant proposals, and joint patents, the expected h-index increases by a factor of 27.37, 23.83, 18.21, and 2.69, respectively (calculated as $e^{(3.345)}-1$, $e^{(3.212)}-1$, $e^{(2.956)}-1$, and $e^{(1.306)}-1$), with the remaining predictor values held constant. That is, it would be expected that a researcher with higher eigenvector centrality score in all networks has a higher h-index score than the other researchers in the College of Engineering. This result was different from the results of Abbasi et al. (2011) which found out that eigenvector centrality had a negative impact on the researcher's citation performance. One reason for this was that the researcher was connected to other researchers who were directly connected to many individual students who already had low collaboration records. However, the results showed that a researcher can be more impactful when the researcher communicates and collaborates with other researchers who are themselves well connected. Abbasi et al. (2011) reported that including demographic information could be useful as moderating variables in the model. Since the log of expected value is modeled as dependent variable in the Poisson regression, coefficients represent the difference in the log of expected value on one level compared with another level for binary or categorical predictors (e.g., demographic attributes) [138]. In almost all models, the difference in the log of the expected h-index were 0.35-0.59 units lower for females than for males, with the rest of the predictor values held constant. That is, females are expected to have 29.6% -55.4% lower h-index than males are in engineering field (calculated as $1\text{-}e^{(-0.35)}$ and $1\text{-} e^{(-0.59)}$). For other demographic variables such as race and department, there were not any overall significant effects on the researchers' citation performance.

Based on the results, *hypothesis 1* is only valid when the social network metrics are obtained from the researchers' collaborative output networks, meaning that the citation performance of a researcher improves to the extent to which the researchers have more distinct connections to other researchers in collaborative output networks than in their communication network. *Hypotheses 2* and *4* can be accepted for all networks. Then, it can be stated that an increase in occupying a central position in both communication and collaborative output networks in terms of the shortness of a researcher's total distance to all other researchers and a researcher's tendency to connect with other researchers who are themselves well-connected will be more advantageous to improve a researcher's citation performance. *Hypothesis 3* only holds for the network of joint publications. This indicates that the citation performance of a researcher improves when the researcher is in the position to broker information and ideas in joint publication relations. *Hypothesis 5* can only be accepted for the networks of joint publications and patents. This means that the citation performance of a researcher improves if there is an increase in the researcher's average number of repeated publications and patents in collaboration with other researchers. *Hypothesis 6* only holds for the network of joint patents. This means that an increasing redundancy of a researcher's joint patent connections to a group of researchers (i.e., inventors in this case) who already generate joint patents together will improve the citation performance of the researcher. *Hypothesis 7* is only valid for the network of joint publications and grant proposals, indicating that a researcher's increasing tendency towards the tight-knit collaborating teams when making publications and submitting grant proposals will improve the researcher's citation performance.

3.5. Discussion

This study is an extension of the study of Abbasi et al. (2011), and it is performed using a richer dataset. Unlike the previous study, this study considers researchers' social network metrics obtained from researchers' multiple collaborative output networks constructed by self-reported data as well as social network metrics obtained from researchers' communication network in a small-scale such as within a college. Additionally, collecting researchers' collaborative output data in a self-reported way provides some indication of whether or not a tie is important in terms of their collaborative research efforts. In other words, the self-reported way of collecting the relations in collaborative outputs permits the researchers to assess both which connection or tie is important to them according to their own perceptions and whether or not reported contact is actually involved in research. Then, the dataset used to construct researchers' collaborative output networks contains richer data since it consists of both in-progress and completed collaborative efforts. This study also considers the local clustering coefficient, i.e., an individual's tendency towards the dense local neighborhoods. It is necessary to consider the local clustering coefficient of a researcher because it is more likely that working in a team, i.e., being in dense-connected cluster leads to higher number of citations [78, 132]. In addition, this study uses h-index instead of g-index because h-index is better to use when researchers within the same field of study are compared [124]. The Poisson regression model was used because h-index is the count data, and the mean and variance of the variable h-index was reasonably close to each other. However, the variance of dependent variable was slightly lower than the mean value of the dependent variable. When this exists, an underdispersion problem occurs. To overcome this problem, and therefore to improve the models, a generalized Poisson regression can be run for all models [143]. Furthermore, Poisson regression is the method of choice for count data, but the h-index is not a pure count variable, but instead a composite index calculated from the rank-frequency distribution. Therefore, there are considerations about how to statistical analyze the h-index, which should be taken into account [144]. The result of Poisson regression bivariate models indicated that unlike the study of the study of Abbasi et al. (2011), eigenvector centrality (i.e., being connected to well-connected researchers) positively impacted the citation performance of the researchers. One reason for this might be that the researchers' connections with students and district connections to other researchers from different colleges are excluded. Furthermore, the previous study found out that closeness and betweenness centralities in the network of joint publications did not significantly impact the citation performance of the researchers, whereas this study detected that their impact was statistically significant and positive.

Table 3.1. Spearman's Rank Correlations

Communication								
	h-index	C_D	C_C	C_B	C_E	*ATS*	*Ef*	*LCC*
h-index	1.000	0.175	0.182	0.127	0.023*	-0.055	0.033	-0.042
$\mathrm{C_D}$		1.000	0.994**	0.968**	0.975**	0.033	0.855**	-0.884**
$\mathrm{C_C}$			1.000	0.956**	0.985**	0.036	0.840**	-0.870**
$\mathrm{C_B}$				1.000	0.912**	-0.019	0.925**	-0.945**
$\mathrm{C_E}$					1.000	0.057	0.776**	-0.809**
ATS						1.000	-0.075	0.065
Ef							1.000	-0.997**
CC								1.000
Joint Publications								
	h-index	C_D	C_C	C_B	C_E	*ATS*	*Ef*	*LCC*
h-index	1.000	0.422**	0.428**	0.241*	0.490**	0.456**	-0.141	0.504**
C_D		1.000	0.912**	0.835**	0.866**	0.393**	-0.124	0.616**
C_C			1.000	0.806**	0.937**	0.402**	-0.092	0.584**
C_B				1.000	0.648**	0.275**	0.110	0.281**
C_E					1.000	0.422**	-0.185	0.685**
ATS						1.000	0.096	0.363**
Ef							1.000	-0.663**
CC								1.000
Joint Grant Proposals								
	h-index	C_D	C_C	C_B	C_E	*ATS*	*Ef*	*LCC*
h-index	1.000	0.309**	0.316**	0.216*	0.336**	0.281**	-0.281**	0.309**
C_D		1.000	0.968**	0.847**	0.875**	0.266**	-0.249*	0.323**
C_C			1.000	0.822**	0.933**	0.267**	-0.237*	0.319**
C_B				1.000	0.664**	0.185	0.086	-0.021
C_E					1.000	0.244*	-0.334**	0.431**
ATS						1.000	-0.173	0.317
Ef							1.000	-0.840**
CC								1.000
Joint Patents								
	h-index	C_D	C_C	C_B	C_E	*ATS*	*Ef*	*LCC*
h-index	1.000	0.288**	0.281**	0.077	-0.033	0.302**	0.304**	0.159
C_D		1.000	0.994**	0.641**	0.622**	0.973**	0.932**	0.532**
C_C			1.000	0.635**	0.658**	0.965**	0.930**	0.523**
C_B				1.000	0.586**	0.483**	0.474**	0.335**
C_E					1.000	0.541**	0.462**	0.511**
ATS						1.000	0.941**	0.517**
Ef							1.000	0.230*
CC								1.000

**<0.01, *<0.05

C_D – Degree Centrality, C_C – Closeness Centrality, C_B – Betweenness Centrality, C_E – Eigenvector Centrality
ATS – Average Tie Strength, *Ef* – Burt's Efficiency Coefficient, *LCC* –Local Clustering Coefficient

Table 3.2. Poisson Regression Results (The h-index as Dependent Variable) for Bivariate Models

	Communication							**Joint Publications**						
Parameter	**Coefficient**							**Coefficient**						
Intercept	**1.051***	-0.051	**1.316***	**0.864***	**1.632***	**0.884***	**1.769***	**0.885***	0.063	**1.210***	**0.930***	**0.734***	**1.154***	**1.204***
C_D	1.077							**10.232***						
C_C		**2.215***							**4.554***					
C_B			1.391							**5.765***				
C_E				**3.345***							**3.212***			
ATS					-0.097							**0.336***		
Ef						0.881							0.218	
LCC							-0.838							**1.200***
Gender [0]	**-0.441***	**-0.444***	**-0.440***	**-0.431***	**-0.430***	**-0.465***	**-0.462***	-0.330	-0.293	**-0.385***	-0.329	**-0.352***	**-0.416***	**-0.590***
Gender [1]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]
Race [1]	0.187	0.180	0.206	0.170	0.197	0.177	0.171	-0.071	-0.036	0.021	-0.104	0.108	0.194	0.167
Race [2]	-0.531	-0.550	-0.509	-0.559	-0.539	-0.592	-0.604	-0.619	-0.471	-0.539	**-0.776***	-0.335	-0.457	-0.390
Race [3]	-0.285	-0.301	-0.256	-0.332	-0.281	-0.301	-0.311	**-0.425***	**-0.411***	-0.433	**-0.383***	-0.138	-0.254	-0.168
Race [4]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]
Department [1]	0.001	-0.004	0.051	-0.031	0.048	0.071	0.067	0.115	0.092	0.157	0.027	0.175	0.100	0.039
Department [2]	-0.126	-0.107	-0.181	-0.049	-0.238	-0.144	-0.139	0.026	-0.077	-0.044	0.184	-0.110	-0.190	-0.300
Department [3]	-0.039	-0.030	-0.080	0.018	-0.081	-0.078	-0.072	0.048	-0.078	-0.042	0.229	-0.022	-0.090	-0.176
Department [4]	0.020	0.022	0.011	0.037	-0.001	0.046	0.047	0.027	-0.098	0.055	0.033	-0.003	0.007	-0.212
Department [5]	-0.624	-0.626	-0.613	**-0.620***	-0.588	-0.602	-0.602	-0.446	-0.643	-0.644	-0.242	-0.495	-0.651	-0.511
Department [6]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]
(Scale)	1[b]				1[b]				1[b]			1[b]		
Likelihood Ratio Chi-Square	33.803	35.203	28.510	39.965	29.491	32.503	33.404	68.194	75.043	44.367	75.879	54.493	29.621	60.344
df	10	10	10	10	10	10	10	10	10	10	10	10	10	10
Sig.	0.000	0.000	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000

*<0.05

[a]Set to zero because this parameter is the base value.

[b]Fixed at the displayed value.

Note: 0= ‘female’, 1= ‘male’ for Gender

1= ‘Asian’ 2= ‘Black’ 3= ‘Hispanic’ 4= ‘White’ for Race

1= ‘CBE’ 2 = ‘CEE’ 3= ‘CSE’ 4= ‘EE’ 5= ‘IMSE’ 6= ‘ME’ for Department

Table 3.2. (Continued)

Parameter	Joint Grant Proposals							Joint Patents						
	Coefficient							Coefficient						
Intercept	**1.038***	0.160	**1.283***	**1.020***	**0.847***	**1.814***	**1.159***	**1.167***	**1.181***	**1.305***	**1.230***	**1.215***	**1.076***	**1.315***
C_D	**3.613***							**15.583***						
C_C		**2.759***							**6.273***					
C_B			3.077							15.993				
C_E				**2.956***							**1.306***			
ATS					0.339							**0.223***		
Ef						-0.652							**0.565***	
LCC							**0.591***							0.368
Gender [0]	**-0.500***	**-0.446***	**-0.445***	**-0.627***	-0.351	**-0.480***	**-0.397***	-0.367	-0.367	**-0.442***	-0.371	**-0.351***	-0.330	-0.416
Gender [1]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]
Race [1]	0.141	0.123	0.184	0.124	0.191	0.251	0.217	0.215	0.186	0.199	0.195	0.276	0.243	0.215
Race [2]	-0.600	-0.633	-0.503	-0.747	-0.576	-0.590	-0.689	-0.645	-0.700	-0.579	-0.533	-0.522	-0.662	-0.505
Race [3]	-.0234	-0.311	-0.247	-0.235	-0.282	-0.175	-0.205	-0.224	-0.229	-0.268	-0.213	-0.333	-0.238	-0.240
Race [4]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]
Department [1]	-0.030	0.079	0.032	-0.197	0.067	-0.063	-0.008	0.022	-0.067	0.055	0.057	-0.106	-0.012	-0.021
Department [2]	-0.103	-0.081	-0.163	-0.0156	-0.204	-0.197	-0.200	-0.066	-0.057	-0.156	-0.098	-0.153	-0.071	-0.182
Department [3]	0.019	0.075	-0.074	0.179	-0.052	-0.133	-0.114	0.015	0.054	-0.054	0.018	-0.187	-0.085	-0.076
Department [4]	0.040	0.045	0.005	0.079	0.001	-0.002	-0.029	-0.104	-0.158	-0.024	-0.109	-0.037	-0.008	-0.017
Department [5]	-0.511	-0.484	-0.622	-0.378	-0.639	-0.526	-0.506	-0.485	**-0.535***	-0.579	-0.546	-0.565	-0.483	-0.610
Department [6]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]
(Scale)	1[b]				1[b]				1[b]			1[b]		
Likelihood Ratio Chi-Square	42.813	46.668	30.224	55.364	35.763	33.979	35.747	47.404	45.676	30.267	42.088	41.195	50.803	30.738
df														
Sig.	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001

*<0.05

[a]Set to zero because this parameter is the base value.

[b]Fixed at the displayed value.

Note: 0= 'female', 1= 'male' for Gender

1= 'Asian' 2= 'Black' 3= 'Hispanic' 4= 'White' for Race

1= 'CBE' 2 = 'CEE' 3= 'CSE' 4= 'EE' 5= 'IMSE' 6= 'ME' for Department

CHAPTER 4: A STRUCTURAL EQUATION MODEL TO TEST THE IMPACT OF RESEARCHERS' INDIVIDUAL INNOVATIVENESS ON THEIR COLLABORATIVE OUTPUTS

4.1. Introduction

Björk and Magnusson (2009) asserted that "innovation can be seen as ideas that have been developed and implemented". When people interact more, the quality of ideas will increase [53]. In addition, working as a group or team stimulates idea generation or ideation [145]. Ideation is a creative process which requires the retrieval of existing knowledge from memory as well as the combination of various aspects of existing knowledge into novel ideas, where an idea is the basic element of thought that can be either concrete or abstract [54]. Due to the associative nature of memory, working in a group and attending to the ideas of others could both spark a good idea from an individual's less accessible area of knowledge and could lead to a novel combination of ideas [54]. Thus, collaboration is necessary for creativity, innovation, and problem solving [54, 55].

From the network perspective, Lovejoy and Sinha (2010) find that individual innovativeness during the ideation phase is accelerated by two properties: 1) an individual's participation in a 'maximal complete sub-graph' or clique, which maximizes the number of parallel conversations, and 2) the knowledge gain of individuals via their conversational churn which means that an individual constantly changes his/her conversational partners through a large set of conversational partners. In addition to these two properties, perceived self-innovativeness should also be considered as an accelerator of the individual innovativeness [57-62]. In the literature, investigating the relationship between researchers' individual innovativeness during ideation phase and their collaborative output is not addressed. This is because the studies in the literature mostly focus on final outputs such as publications and citations due to the major limitation of collecting information with regard to researchers' interaction in the early stage of their collaborative activities. The findings of Lovejoy and Sinha (2010) can be used to test to what extent researchers' individual innovativeness impacts the number of their collaborative outputs (joint publications, grant proposals, and patents). Since knowledge creation is an important step which supports idea generation [63] and the strength of an interpersonal connection impacts how easily the created knowledge can be transferred to other individuals [64-67], it is also important to consider the tie strength of a researcher to other conversational partners while investigating the relationship between researchers' individual innovativeness and their collaborative outputs. Thus, this chapter seeks an answer for the following question: *what is the impact of researchers' individual innovativeness (as determined by the specific indicators obtained from their communication network) on the volume of their collaborative outputs taking into account the tie strength of a researcher to other conversational partners?*

4.2. Literature Review and Hypotheses

4.2.1. The Effect of Individual Innovativeness (Iinnov) on Researchers' Collaborative Outputs (CO)

Communication between individuals enhances innovation because they acquire knowledge due to exposure to different and diverse ideas from others [146-149]. Similarly, Rogers (1995) purported that "we must understand the nature of networks if we are to comprehend the diffusion of innovations fully" because communication involves information exchange in interpersonal networks whereby individuals accumulates knowledge. Using the network of interpersonal interactions, increasing current knowledge level by incorporating new inputs from others and implementing new ideas from these inputs is an important source of individual innovativeness for researchers [53, 151]. Thus, acquiring ideas from the repositories of different knowledge sets, selecting and adopting the most useful ones, and recombining and transforming these acquired ideas in a novel way are the key steps to be able to innovate. Coleman (1988) viewed the social cohesion engendered by a closed network structure as the source of willingness to transfer knowledge between individuals because this type of network structure reduced the risk of knowledge exchanges due to the fact that group norms and rules facilitated cooperation between individuals by constraining exploitive behavior [56, 67, 153]. Additionally, individuals should constantly change their interaction partners to be exposed to different ideas, thereby increasing their current knowledge levels and they should utilize their innate innovativeness. This study proposes that individual innovativeness during the ideation phase is accelerated by three properties each of which is discussed below in detail.

1. *Researcher's rate of participation in 'complete graph(s)':* Network structure facilitates the creation of innovation [154]. To understand this network structure effect, two competing network views in social capital theory, the network closure effect and structural holes effect, can be visited [155-157]. First, Coleman (1988) highlighted that networks with closure in which every individual is connected, i.e. dense sub-groups is the primary source of the creation of innovation due to the fact that individuals are more likely to share tacit knowledge[1] [157]. Second, Burt (1992) purposed that networks with weak network

[1] Knowledge is divided into two types: explicit and tacit [190]. Explicit or codified knowledge is easily transmittable to another person by either writing it down or articulating it, e.g., user manuals, documents, whereas tacit or

architecture or containing 'structural holes' are also the source of the creation of innovation because individuals who locate themselves to close these structural holes can function as a bridging or bonding actor and combine both novel ideas and non-redundant information which flow through different clusters [153, 156, 158-160]. In Coleman's view, the presence of cohesive ties (i.e., network closure) promotes a normative environment which helps create trust and cooperation and strengthen the solidarity between individuals [153, 154]. A maximal complete sub-graph, or a clique (see Figure 4.1), is the maximum number of actors who have all possible ties present among themselves [56]. Referring to Coleman's network closure definition, a clique-type of network structures can be used to measure the degree of cohesiveness between individuals. Several studies highlighted that there was a positive impact of the clique-type of network structures on individuals' innovativeness [146, 148, 161, 162]. One recent study by Lovejoy and Sinha (2010) found that individual innovativeness during the ideation phase was accelerated by the clique-type of network structures (called just 'complete graphs' in their study).

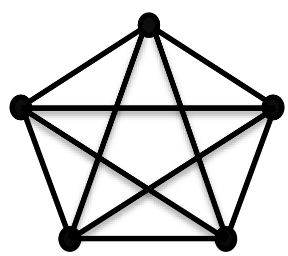

Figure 4.1. 'A Maximal Complete Sub-graph' Consisting of 5 Actors

2. *Researchers' knowledge gain (KG) via conversational churn*: Innovation depends on the availability of knowledge [163]. Knowledge is defined as "the state of knowing and understanding" and knowledge management involves building and managing knowledge

noncodified knowledge is difficult to transfer by either writing it down or articulating it, and it requires direct experience, e.g., using an complex equipment and ability to speak languages [190].

stocks [164]. Bozeman and Rogers (2002) proposed a churn model that is a process during which individual researchers accumulate or gain knowledge, thus enhance their capabilities, as a result of interactions within networks (also called knowledge value collective) that is a set of individuals connected by their uses of a body of scientific and technical knowledge. Lovejoy and Sinha (2010) evaluated the churn model effect by performing a network simulation in which the knowledge of each individual is represented by binary strings consisting of 1s and 0s and altered through an individual's interaction (or conversational exchanges) with others. Thus, the individual reaches to the "great idea" or "aha moment" when 0s in his/her knowledge string are converted to all 1s. They found that individual innovativeness during the ideation phase was accelerated by two properties. The first one is an individuals' participation in a 'maximal complete sub-graph' type network structure (a cohesive subunit) which maximizes the number of parallel conversations. The second one is the KG of individuals via their conversational churn which is defined as an individual's constantly changing of his/her conversational partners through a large set of conversational partners. This study proposes a formula which calculates an individual's KG via conversational churn using empirical data. The formula is shown in Eq. (4.1):

$$KG = \sum_{i=1}^{6} n_i + \sum_{i=1}^{6} f(t_i) C_i n_i \tag{4.1}$$

$$f(t) = \frac{2^{\alpha t} - 1}{\max(2^{\alpha t} - 1)} \tag{4.2}$$

where i refers to the levels (or periods) in the Likert scale (see Q2 in the Appendix A). Since 6 Likert scale [once a day(6), once a week(5), once every two week(4), once a month(3), once every two months(2), once every three months(1)] is used in the study, the total number of periods is 6. n_i indicates the total number of conversational partners at each specific level. C_i is the number of conversations a researcher has during a period. For example, in a year, a researcher can have 260 daily conversations (considering business days only), 52 weekly conversations, 26 biweekly conversations, 12 conversations once a month, 6 conversations once every two months, and 4 conversations once every three months. *f(t)* refers to the knowledge growth function by which a researcher accumulates knowledge on a daily basis. As shown in Eq. (4.2), in this study, 2 was chosen as the base in the function of *f(t)* and α determines the shape of the parabola capturing the growth rate of knowledge. This study used 0.05 for α. By incorporating the denominator into *f(t),* the maximum value of *f(t)* a researcher's knowledge can grow is 1, which is during the period of three months (see Figure 4.2). Eq. (4.1) has two parts. The first part, $\sum_{i=1}^{6} n_i$, computes the total knowledge value a researcher extracts from all of his/her reported conversational partners. For example, when a researcher meets with his/her conversational partner to exchange information on day 0 (a sort of an initial state) assuming that they have not done so for a while (this study assumed for three months) the researcher can obtain the maximum value of knowledge from the conversation, which is 1. Thus, the researcher can obtain the value of 1 from each of his/her conversational partners. The second part, $\sum_{i}^{6} f(t_i) C_i n_i$, computes how much total knowledge gain a researcher can obtain from the conversations with his/her partner if he/she meets with the same researcher the next day, a week later, two week later, a month later, two months later, or three months later. This part takes into account the fact that if the researcher meets with the same partner next day it is less likely that they exchange new information, but if they wait more it is more likely that they exchange new information. Therefore, KG of the researcher if he/she waits for one day is less than KG of researcher if he/she waits for a week, and KG of the researcher if he/she waits for a week is less than KG of the researcher if he/she waits for two weeks, and so on. Using the values of 0.05 for α and 2 for the base in f(t) ensures that the value of knowledge growth for a researcher are moderately kept low for the interactions: once a day, once a week, and once every two week, but maximally high for the interactions: once a month, once every two months, and once every three months.

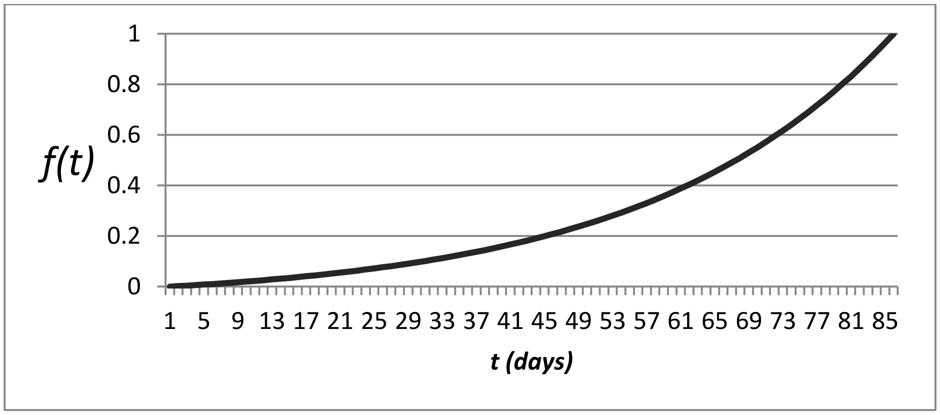

Figure 4.2. Knowledge Growth Function

3. *The perceived self-innovativeness of researchers*: An individual's personality or innate characteristics contribute to his/her innovativeness [57-62]. Rogers (1995) proposed that individuals were characterized as innovative as long as they early adopt an innovation. However, Midgley and Dowling (1978) criticized this notion in a way that innovativeness could not be dependent on observable phenomena such as the time of adoption, rather it existed only "in the mind of the investigator and at a higher level of abstraction". Flynn and Goldsmith (1993) also defended that individual innovativeness should be measureable from a global perspective called global innovativeness that is "a personality dimension that cut across the span of human behavior". By using a 20-item questionnaire (see the Appendix A), Hurt et al. (1977) first attempted to assess an individual's innovativeness as his/her personality trait which was defined as "perceived willingness to change". This study used the questionnaire developed by Hurt et al. (1977) to measure the extent to which a researcher's innate characteristics contributes to his/her innovativeness.

This study investigates the impact of researchers' individual innovativeness, as determined by the specific indicators obtained from their interactions in the early stage of their collaborative network activities, on the number of collaborative outputs that can be considered as a measure of innovative output produced. Then, the following hypothesis is purposed:

*Hypothesis 1***:** There is a positive impact of researchers' individual innovativeness on the volume of researchers' collaborative outputs.

4.2.2. Tie Strength of an Individual to Other Conversational Partners (TS)

Knowledge creation is an important step which supports idea generation [63]. Informal interpersonal connections between individuals play a critical role in knowledge creation and transfer [67]. Additionally, the strength of an interpersonal connection impacts the ease with which created knowledge is transferred to other individuals [64-67]. In the literature, both strong ties and weak ties, two views of tie strength, have been purported to enhance an individual's knowledge [166]. Strong ties between individuals promote information flow about activities within an organizational subsystem, while weak ties between individuals promote information flow about activities outside an organizational subsystem [167, 168]. Hansen (1999) made a similar point which was that the transfer of tacit knowledge is easier between individuals who have strong ties, whereas the transfer of explicit knowledge is easier between individuals who have weak ties. Krackhardt (1992) showed that strong ties are important since they generate trust. Therefore, strong ties lead to greater knowledge exchange between individuals by ensuring that knowledge seekers sufficiently understand each other [64, 65, 166, 169]. Strong ties tend to bond similar individuals to each other and cluster them together; hence, individuals are all connected to each other. Therefore, information obtained via strong ties is more likely to be redundant and this hinders a network from becoming a channel for innovation [65, 169]. In contrast, weak ties behave like local bridges and reach out to nonredundant information from the disparate parts of the system [70, 166, 169]. Then, weak ties combine the ideas from different sources with fewer concerns regarding social conformity, which positively influences individuals toward their innovative propensities [150, 170]. From another viewpoint, Rost (2011) demonstrated that individuals with strong ties, but embedded in weak network structures (structural holes or a peripheral network position) came up with the most innovative solutions. Granovetter (1973) proposed that tie strength was "a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding) and the reciprocal services which characterize the tie" [71]. This study uses the first three of these four indicators (or dimensions). The amount of time spent was measured by asking the question (Q1) "how frequently do you exchange conversations or ideas?" and was called 'frequency' [66, 71]. 'Closeness' is used as a measure of the emotional intensity of a relationship, and the question (Q2) "how close is your relationship between you and your conversational partner?" was asked to assess this dimension [66, 71]. Respondents were asked the question (Q3) "how often do you discuss your work or home personal problems with your conversational partner?" which measures the extent of mutual confiding (intimacy) between individuals [71, 171, 172]. Based on the discussion made so far, it is also important to consider TS and to test the impact of TS on their individual innovativeness, the volume of their collaborative outputs, and the relationship between researchers' individual innovativeness and the volume of their collaborative outputs. Therefore, this study asserts the following three hypotheses:

Hypothesis 2 & *Hypothesis 3*: There is a non-zero impact of TS on both researchers' individual innovativeness and the volume of researchers' collaborative output.

Hypothesis 4: There is a non-zero impact of TS on the relationship between researchers' individual innovativeness and the volume of researchers' collaborative outputs.

4.3. Method

4.3.1. Constructing Dataset for Statistical Model

For 100 tenured/tenured-track faculty members, 9 variables are available. That is, a 100x9 data matrix was compiled. The variables included in the dataset are researchers' rate of participation in 'complete graph(s)'; researchers' knowledge gain via their conversational churn; the perceived self-innovativeness score of researchers; the number of joint publications, grant proposals, and patents; and researchers' total scores for the frequency of communication with other researchers and the strength of closeness and intimacy in their communication ties with other researchers. Researchers' rate of participation in 'complete graph(s)' was computed from an actor-by-actor clique co-membership matrix using UCINET version 6.308. The perceived self-innovativeness score of researchers was measured by employing a 20-item questionnaire and the score received for each researcher was computed [87]. The number of joint publications, grant proposals, and patents was calculated by averaging the rows or columns of data matrixes constructed from collaborative output tie information provided by participants (see section 2.3.3). For a researcher, three dimensions of tie strength (i.e., 'frequency', 'closeness', and 'intimacy') were recorded in three 100x100 data matrixes constructed via three questions answered by the researchers in the survey. Table 4.2 shows three cases that were encountered in the data matrixes.

Total scores for three dimensions of tie strength should be calculated for each researcher. The calculation was done in two steps. First, three data matrixes constructed for each TS indicator were converted into new data matrixes by a method used in the study of Mathews et al.

(1998). The method was revised and applied to three cases in a way as shown in Table 4.3. Second, either each column or each row of these converted data matrixes was summed in order to obtain the total score for each TS indicator for a researcher.

4.3.2. Statistical Model

The observable variables are assigned to 3 latent variables-LVs (or constructs) as shown in Table 4.1. The partial least squares (PLS) path modeling is used to run 3 different path models using these 3 LVs. The three path models each of which test the above-mentioned three proposed hypotheses were run by the SmartPLS computer package using the bootstrap resampling procedure, a non-parametric method, to test the significance of LV loadings and paths between LVs.

Table 4.1. Assignment of Observable Variables to Latent Variables

Latent Variables	**Observable Variables**		
Tie strength of an individual to others (TS)	Frequency	Closeness	Intimacy
Collaborative Outputs (CO)	The number of joint publications	The number of joint grant proposals	The number of joint patents
Individual Innovativeness (Iinnov)	Researchers' rate of participation in 'complete graph(s)'	Researchers' knowledge gain via their conversational churn	The perceived self-innovativeness score of researchers

4.4. Results

4.4.1. Partial Least Squares (PLS) Path Models

Structural Equation Modeling (SEM) is a statistical technique that enables the researchers to construct unobservable variables measured by indicators, and to test and estimate the casual relationships between those LVs [174]. There are two approaches to estimate those relationships: the covariance-based approach and the variance-based (or PLS) approach. The former uses maximum likelihood estimation (MLE) to minimize the difference between the sample covariance matrix and the covariance matrix predicted by the proposed theoretical model and MLE assumes that the joint distribution of variables in the model follows a multivariate normal distribution, whereas the later maximizes the explanation of variance by estimating the partial model relationships in an iterative sequence of ordinary least squares (OLS) regressions [175, 176]. The PLS approach originally developed by Wold (1985) offers several minimal requirements of restrictive assumptions compared to the covariance-based approach that can primarily be attributed to Karl Jöreskog [178] who introduced the particular formulation which is the LISREL model [176].

The PLS path modeling is a “soft” structural equation modeling (SEM) technique because it has very few distribution assumptions and few cases can suffice, unlike the “hard” SEM technique, which requires heavy distribution assumptions and several hundreds of cases [179]. The PLS path modeling is more suitable for a theoretical framework that is not fully crystallized, a complex model that has a large number of indicators and LVs, a model that has LVs constructed in a formative way (i.e., arrows from indicators are directed to LVs), and data that does not satisfy the assumptions of multivariate normality, independence and large sample size [180-182]. This study uses social network metrics such as researchers’ rate of participation in ‘complete graph(s)’ as variables in the model, meaning that the assumption of independence of observations of each other is violated for those variables. Therefore, running the PLS path modeling over the dataset used in this study is more suitable. The model validation in PLS path models is an attempt to assess whether two stages of a model (the measurement model and the structural model) fulfill the quality criteria for empirical work [175]. Therefore, the path models must be analyzed and interpreted for those two stages [175, 176, 182, 183].

The measurement (or outer) model is defined as the relations between indicators and LVs, and it is evaluated in the first stage. It can be constructed as either reflective way (outwards directed) or formative way (inwards directed) based on the unidimensionality or homogeneity of the block of indicators. All blocks are considered homogenous, if Cronbach's alpha is higher than 0.7 [179, 184]. In this study, Cronbach's alphas in all models were very close to this threshold value, indicating that selecting the reflective way was appropriate. In a reflective model, the relationship between each indicator, *p*, and its LV, ξ_q, is shown by a simple linear regression in Eq. (4.3a):

$$x_{pq} = w_{p0} + w_{pq}\xi_q + \varepsilon_{pq} \tag{4.3a}$$

$$E(x_{pq} \mid \xi_q) = w_{p0} + w_{pq}\xi_q \tag{4.3b}$$

where w_{pq} is the loading (or weight) associated to the *p*-th indicator for *q*-th LV and ε_{pq} is the related error term [184]. The assumption for this model is that the error term ε_{pq} has a zero mean and is uncorrelated with LV, ξ_q. Then, the Eq. (4.3a) is reduced to the Eq. (4.3b).

The Structural (or inner) model is defined as the relations between LVs and is evaluated in the second stage. Each LV, $\xi_{q'}$, is regressed on other Q LVs, ξ_q, shown as.

$$\xi_{q'} = \beta_{q'0} + \sum_{q=1}^{Q} \beta_{q'q} * \xi_q + \nu_q \tag{4.4a}$$

$$E(\xi_{q'} \mid \xi_q) = \beta_{q'0} + \sum_{q=1}^{Q} \beta_{q'q} * \xi_q \tag{4.4b}$$

where $\beta_{q'q}$ are regression coefficients (or inner weights) between LVs and ν_q is the error term related to $\xi_{q'}$ [179, 184]. Since the assumption is the error term ν_q which has zero mean and no correlations with LVs ξ_q in the model, the Eq. (4.4a) is reduced to the Eq. (4.4b) [182]. PLS algorithm first assigns arbitrary initial outer weights and estimates LVs using these initial weights. After the estimation, ordinary least squares (OLS) regression is run between estimated LVs to find the inner weights, and the previously estimated LVs are updated based on these inner weights. In other words, the inner weights are estimated using the calculated LV scores in accordance with the specified network of structural relations. The estimation of the outer weights is iterated until the convergence is observed by means of the alternation of the *outer* and the *inner* estimation steps [184]. The estimation of outer weights from the updated LV estimates is done using either individual OLS regression per indicator if outer model is a reflective construct or a multiple regression if outer model is a formative construct. The estimation procedure is called partial because it solves block one at a time via alternating the single and multiple linear regressions [184]. During the step where OLS regression is performed between LVs, PLS regression can be used if LVs are highly correlated [184]. The PLS path modeling was performed using the SmartPLS package version 2.0.M3, and the results for Model 1, 2, and 3 are shown in Figure 4.3, 4.4, and 4.5 and Table 4.4a&b, 4.5a&b, and 4.6a&b. The next section discusses each stage in detail.

4.4.2. Analysis of Partial Lease Squares (PLS) Models

4.4.2.1. Assessment of Measurement Models

A measurement model is assessed with regard to the reliability and validity of the LVs in the model. Once the outer model shows the evidence of sufficient reliability and validity, it will be more meaningful to evaluate the inner path model estimates [182]. The measurement models were assessed by the following criteria summed up by Urbach and Ahleman (2010).

1. *Internal consistency reliability* (ICR): There are two criteria to assess ICR: a Cronbach’s alpha (α) measure and a composite reliability measure. Cronbach's α is a measure of internal

consistency, and it is used to measure how closely related a set of items are as a group [185]. The composite reliability (CR) measure relaxes the Cronbach's α assumption that all scale items are equally related to the attendant LV [175]. Otherwise, Cronbach's α will tend to underestimate the ICR of LVs. Both of these measures were close and above the threshold value of 0.70, which indicated the adequate internal consistency [175].

2. *Indicator reliability* (IR): A LV should explain a substantial part of each indicator's variance, which is usually at least 50% [182]. Then, a variable and set of variables will be consistent about what it really intends to measure. To assess IR, indicator loadings should be both statistically significant at the 0.05 significance level and higher than 0.7 (square root of 50%) [175, 186]. The significance of both LV loadings and the associations between LVs is determined via the bootstrap procedure that is a resampling method [187]. In this procedure, the proposed model is run several times (this study ran 1000 times) using repeated random samples of each items in order to construct a distribution for each association. Thus, where the original value falls in this distribution is investigated by calculating a t-value statistics (or related p-value). While running bootstrap resampling procedure in the SmartPLS, the option of 'individual changes' for sign changes was selected [182]. All LV loadings in three models were significant at the 0.05 level and they were close to or mostly higher than the threshold value of 0.70.

3. *Convergent validity* (CV): A set of indicators representing the same underlying construct should converge or demonstrate a unidimensionality compared to the indicators representing other constructs. To assess CV, average variance extracted (AVE) is commonly used, measuring the amount of variance that LV captures from its indicators relative to the amount due to the measurement error [188]. AVEs for all LVs across all models were all above 0.50

(threshold value), which indicated sufficient CV. This should be interpreted that all LVs were able to explain more than half of the variance of its indicators on average [182].

4. *Discriminant validity* (DV): Any single construct (or LV) should be different from the other constructs in a proposed model. In other words, two conceptually different constructs should exhibit sufficient difference [182]. There are two commonly applied criteria to assess DV: the cross-loadings and The Fornell–Larcker criterion. In the cross-loading criterion, the loadings of each LV are expected to be higher than all of its cross-loadings with other LVs in the proposed model [182, 186]. Then, it can be inferred that there is a sufficient difference between constructs. The Fornell–Larcker criterion requires that a LV has to share more variance with its assigned indicators than with any indicators of other LVs [182, 186]. Then, according to the Fornell–Larcker criterion, DV is assessed by that the AVE of each LV should be greater than squared correlations with other LVs [182]. With cross-loadings criteria, the LVs in all models indicated a moderate DV. With Fornell–Larcker criterion, a square root of AVE for an LV was compared to the LV's squared correlation with any other LV and it was again observed that the LVs in all models indicated a moderate DV.

4.4.2.2. Assessment of Structural Models

Exogenous LVs are the constructs that do not have any predecessors or only have arrows originating from them in the structural model, whereas endogenous LVs are the constructs which has one or more arrows leading into it [176]. A structural model is assessed to determine the significance of the inner paths or hypothesized paths and its explanatory power using the amount of variance accounted for by the endogenous constructs [189]. The structural models were assessed by the following criteria:

1. *Coefficient of determination*: R-square (also called coefficient of determination) measures the amount of variance in the construct that is explained by the model [183]. In other words, it measures the relationship of a construct's explained variance to its total variance. Chin (1998) considers R-square values of 0.67, 0.33, and 0.19 in PLS path model as substantial, moderate, weak. As seen from all three models, R-square values were either moderate or substantial. For example, R-square value in Model 1 was 0.415, meaning that approximately 42% of variance in construct CO was explained by the exogenous construct Iinnov.
2. *Evaluation of path coefficients*: The individual path coefficient of the PLS structural model is interpreted as standardized beta coefficients of ordinary least squares regressions [182, 189]. The path coefficients are tested by assessing the direction, strength, and the level of significance (the bootstrap resampling method with 1000 resamples was used to test the significance). Testing the path coefficients provides a partial empirical validation of theoretically assumed relationships (i.e., hypotheses) between constructs [182]. Path coefficients showing insignificance and signs contrary to hypothesized direction do not support a prior hypothesis, whereas paths showing significance and a sign fitting empirically support the casual relationship [189]. The values for the path coefficients in PLS models are given in the standardized form (i.e. between 0 and 1). The path coefficients corresponding to 4 hypotheses are statistically significant in all models. The model 1 corresponding to Hypothesis 1 presents high and positive value of the path coefficient, indicating that for one unit change in researchers' individual innovativeness, collaborative outputs increases by 0.644. Then, this indicates that the conversion rate of researchers' ideas into the number of their collaborative outputs is high in the college of engineering. Based on the definition of tacit and explicit knowledge [190], the constructs Iinnov and CO can be considered as tacit

and explicit knowledge, respectively. Then, testing *hypothesis 1* attempts to fill the gap in knowledge creation literature which is the process of the conversion of tacit knowledge into explicit knowledge (also called ‘externalization’) [191-193]. The model 2 corresponding to *hypothesis 2* and *3* tests this conversion in the presence of researchers’ strength of interpersonal connections. It can be seen that there is a higher and positive increase in the conversion rate when the construct TS directly impacts the two constructs Iinnov and CO. Therefore, *hypothesis 2* confirms previous literature that the transfer of tacit knowledge is easier between individuals who have strong ties [66]. The result of *hypothesis 3* presents a moderately low and negative direct impact of tie strength of an individual to others and fits the theory of ‘strength of weak ties’ proposed by Granovetter (1973).This indicates that the weaker ties researchers have with others in the early stages of their collaborative activities the more they have the final collaborative outputs. The result also matches up with the finding of Hansen (1999) which was that the transfer of explicit knowledge was easier between individuals who have weak ties. The model 3 corresponding to *hypothesis 4* tests the moderating effect of researchers’ strength of interpersonal connections in the impact of researchers’ individual innovativeness on their collaborative outputs. In PLS, the moderating effect is the interaction term which is built by the products of each indicator of the independent latent variable Iinnov with each indicator of the moderator variable TS [194]. From model 3, it can be seen that there is a low and negative moderating effect of TS, indicating that the theory of ‘strength of weak ties’ rules the process of the conversion of tacit knowledge into explicit knowledge.

3. *Redundancy index* (RI) *or Redundancy*: RI is a measure of the quality of the structural model for each endogenous block by taking the measurement model into account [179]. In other

words, RI measures the portion of variability of the manifest variables connected to the endogenous LV explained by the LVs directly predicting the same endogenous LV [184]. It is the measure of the quality of structural model for each endogenous construct and calculated by multiplying the average communality of a construct (i.e., AVE) by R-square of the same construct [179]. The following redundancy assessment scale was derived by substituting the minimum average of AVE of 0.50 as suggested by Fornell and Larcker 1981 and the Chin (1998)'s proposed scale for R-squares values at substantial, moderate, and weak level in the equation defining redundancy (redundancy=communality*R-square); $Redundacy_{substantial}$= 0.34, $Redundacy_{moderate}$=0.17, and $Redundacy_{weak}$=0.10. Redundancy in all of the three models ranged from moderate to substantial.

4. *Cross-Validated (Communality and Redundancy) index*: Besides checking the magnitude of R-squares to assess the predictive relevance, the predictive sample reuse technique, called the Stone-Geisser test criterion (or Q^2), can also be used [183]. The Q^2 test statistics is a jackknife version of the R-square statistics [179]. Chin (1998) stated that Q^2 statistics is a measure of how well observed values are reconstructed by the model and its parameter estimates. Calculation of Q^2 involves 1) omitting (or blindfolding) one case at a time, 2) re-estimating the model parameters by using the remaining cases, and 3) predicting the omitted case values based on the remaining parameters [179]. Q^2 statistics can be obtained through two ways: *cross-validated communality* Q^2, also called H^2, in which prediction of the data points is made by the underlying LV score, *cross-validated redundancy* Q^2, also called F^2 in which prediction is made by those LVs that predict the block in question [179]. $Q^2>0$ implies the model has predictive relevance whereas $Q^2<0$ represents a lack of predictive relevance. For three models, blindfolding procedure has been performed using G=7 (G is the omission

distance. For further discussion of G, please see Tenenhaus et al. (2005) p.175). The value of Q^2 was greater than 0 in all of the three models, indicating that all models has predictive relevance.

5. *Goodness of fit Index* (GoF): GoF index evaluates the model performance by taking both measurement and structural model into consideration and thus offer a single measure for the overall prediction performance of the model [184]. GoF index is calculated by the following formula: GoF=$\sqrt{\overline{AVE} \times \overline{R^2}}$. Threshold values were calculated by plugging a cut-off value of 0.5 for communality and the cut-off values for R-square proposed by Chin (1998) into the formula. The baseline values for $GoF_{substantial}$, $GoF_{moderate}$, and GoF_{weak} were obtained 0.58, 0.41, and 0.31. Only GoF index for peers has a fit for the weak level. All of the three models indicated the moderate and weak GoF values, concluding that the models had an adequate explaining power in comparison with baseline values.

4.5. Discussion

This study seeks to contribute to the informetrics literature by proposing a model that investigates the relationship between researchers' individual innovativeness and their collaborative output. PLS path modeling does not require the assumptions of multivariate normality, independence of observations, and large sample size. This study used social network metrics such as researchers' rate of participation in 'complete graph(s)' as variables in the model, meaning that the assumption of independence of observations is violated, then running the PLS path modeling over the dataset used in this study is more suitable. A formula, which measures an individual's KG via conversational churn using empirical data, was proposed. Two properties accelerating individual innovativeness which was found in the study of Lovejoy and Sinha (2010), 1) participation in a 'maximal complete sub-graph' or clique and 2) KG via conversational churn, was empirically tested and found that both of these properties were statistically significant.

Table 4.2. The Cases Observed in Matrixes

Case 1 (Both scored each other)		
	Researcher	Researcher’s partner
Researcher		**X**
Researcher’s partner	**X**	
Case 2 (Only a researcher scored his/her partner)		
	Researcher	Researcher’s partner
Researcher		**X**
Researcher’s partner		
Case 3 (Only a researcher’s partner scored the researcher)		
	Researcher	Researcher’s partner
Researcher		
Researcher’s partner	**X**	

Table 4.3. A Method to Convert the Data Matrixes for TS Indicators

Case 1 (Both scored each other)

Both a researcher and his/her partner get scored **3** in case

- Both the researcher’s score for his/her partner is *greater* than the researcher’s mean score for all of his/her communication partners
- and his/her partner’s score for the researcher is *greater* than the partner’s mean score for all of his/her communication partners

Both a researcher and his/her partner get scored **2** in case

- Both the researcher’s score for his/her partner is *greater* than the researcher’s mean score for all of his/her communication partners
- and his/her partner’s for the researcher is *lower* than mean score for the partner’s mean score for all of his/her communication partners

Or

- Both the researcher’s score for his/her partner is *lower* than the researcher’s mean score for all of his/her communication partners
- and his/her partner’s score for the researcher is *greater* than the partner’s mean score for all of his/her communication partners

Both a researcher and his/her partner get scored **1** in case

- Both the researcher’s score for his/her partner is *lower* than the researcher’s mean score for all of his/her communication partners
- and his/her partner’s score for the researcher is *lower* than the partner’s mean score for all of his/her communication partners

Case 2 (Only a researcher scored his/her partner)

Both a researcher and his/her partner gets scored **2** in case

- The researcher’s score for his/her partner is *greater* than the researcher’s mean score for all of his/her communication partners

Both a researcher and his/her partner gets scored **1** in case

- The researcher’s score for his/her partner is *lower* than the researcher’s mean score for all of his/her communication partners

Case 3 (Only a researcher’s partner scored the researcher)

Both a researcher and his/her partner gets scored **2** in case

- His/her partner’s score for the researcher is *greater* than the partner’s mean score for all of his/her communication partners

Both a researcher and his/her partner gets scored **1** in case

- His/her partner’s score for the researcher is *lower* than the partner’s mean score for all of his/her communication partners

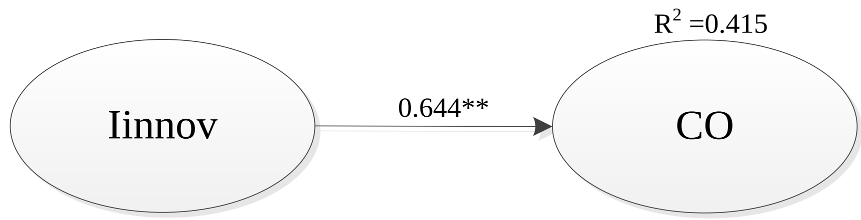

Figure 4.3. Illustration of Model 1
0.05<**, 0.1<*

Table 4.4a. LV Loadings and Assessment of Measurement Model for Model 1

	Individual Innovativeness (Iinnov)	Collaborative Outputs (CO)
Cpart	**0.837**	0.458
Kgain	**0.863**	0.630
Sinnov	**0.583**	0.348
Publication	0.510	**0.870**
Grant	0.659	**0.863**
Patent	0.314	**0.696**
Cronbach's α	0.656	0.756
CR	0.811	0.853
AVE	0.595	0.662
Sqrt(AVE)	0.771	0.814
LV correlations	0.644 (Iinnov-CO)	

Cpart – Researchers' rate of participation in 'complete graph(s)'
Kgain – Researchers' knowledge gain via their conversational churn
Sinnov – The perceived self-innovativeness score of researchers
Publication – The number of joint publications
Grant – The number of joint grant proposals
Patent – The number of joint patents

Table 4.4b. Assessment of Structural Model for Model 1

	Redundancy	H^2	F^2	GoF
Iinnov	0.000	0.238	0.238	0.361
CO	0.275	0.335	0.244	

H^2 – cross-validated communality
F^2 – cross-validated redundancy
GoF – goodness of fit index

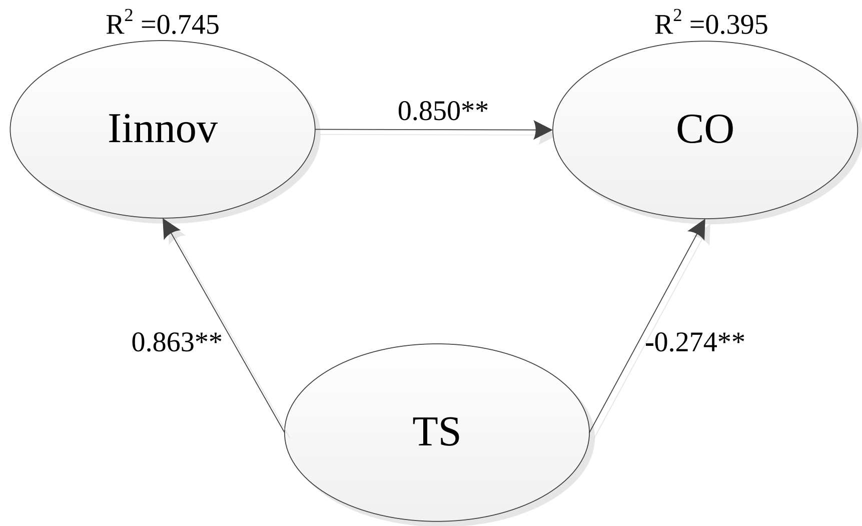

Figure 4.4. Illustration of Model 2
0.05<**, 0.1<*

Table 4.5a. LV Loadings and Assessment of Measurement Model for Model 2

	Individual Innovativeness (Iinnov)	Collaborative Outputs (CO)	Tie Strength (TS)
Cpart	**0.885**	0.448	0.904
Kgain	**0.802**	0.627	0.541
Sinnov	**0.606**	0.348	0.494
Publication	0.477	**0.877**	0.287
Grant	0.656	**0.851**	0.605
Patent	0.277	**0.709**	0.104
Frequency	0.849	0.435	**0.986**
Closeness	0.864	0.454	**0.987**
Intimacy	0.838	0.469	**0.982**
Cronbach's α	0.660	0.760	0.990
CR	0.813	0.856	0.990
AVE	0.600	0.670	0.970
Sqrt(AVE)	0.775	0.819	0.985
LV correlations	0.613 (Iinnov-CO)		
	0.863 (Iinnov-TS)		
	0.459 (CO-TS)		

Frequency – Frequency of communication between researchers
Closeness – The strength of emotional intensity
Intimacy – The strength of mutual confiding

Table 4.5b. Assessment of Structural Model for Model 2

	Redundancy	H^2	F^2	GoF
Iinnov	0.447	0.241	0.427	0.533
CO	0.265	0.331	0.240	
TS	0.000	0.867	0.867	

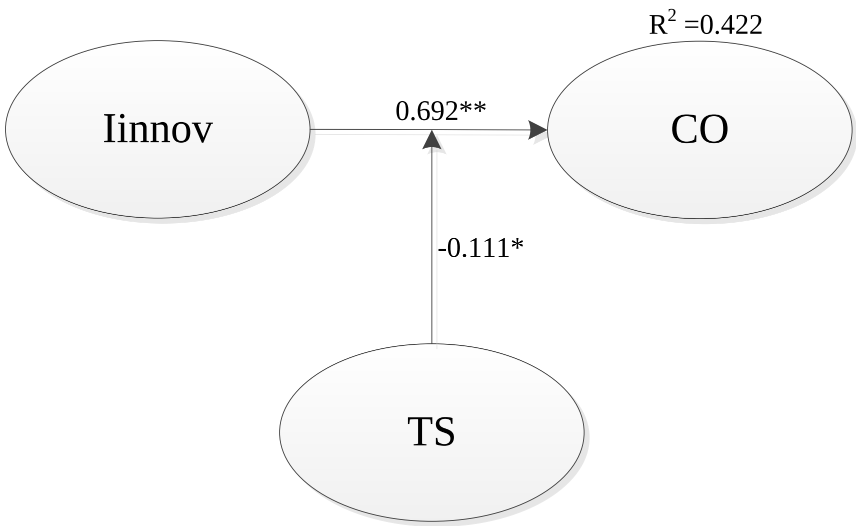

Figure 4.5. Illustration of Model 3
0.05<**, 0.1<*

Table 4.6a. LV Loadings and Assessment of Measurement Model for Model 3

	Individual Innovativeness (Iinnov)	Collaborative Outputs (CO)
Cpart	**0.835**	0.453
Kgain	**0.863**	0.629
Sinnov	**0.584**	0.348
Publication	0.510	**0.874**
Grant	0.658	**0.857**
Patent	0.315	**0.701**
Cronbach's α	0.656	0.756
CR	0.811	0.854
AVE	0.595	0.664
Sqrt(AVE)	0.771	0.815
LV correlations	0.642 (Iinnov-CO)	

Table 4.6b. Assessment of Structural Model for Model 3

	Redundancy	H^2	F^2	GoF
Iinnov	0.000	0.231	0.231	0.287
CO	0.280	0.337	0.263	

CHAPTER 5: CONCLUSION

The findings of these three studies offer several implications for college and university administrations as well as for policy makers in their attempt to prosper the collaborative relationships between researchers. With the results of this study, the college administration is informed regarding the extent that the social cohesion formed by interpersonal ties impacts on or drives the collaboration activity that resulted in collaborative outputs. In addition, the results help the college administration to find out the collaborative tendency of each researcher in different networks, and prolific researchers and departments determined by social network metrics (e.g., centrality metrics for individuals and groups) can be rewarded. Using the results, the college administration also finds out to what degree a department is more inclined to form external ties in its collaboration activity. Collaboration is related to many types of shared attributes [16, 30]. Then, the results of this study also have the potential to identify connections of members from underrepresented groups (e.g., female researchers and black/African American researchers) in their networks in order to establish research collaborations between them and other members, in case connections to members of underrepresented groups are insufficient (or non-existent).

This study has the potential to be generalized and applied other colleges and disciplines, and even the university as a whole. Within a university, structural properties of these four networks across different colleges can be compared in order to help university administration to understand the nature of collaboration of each college and interdisciplinary relations. Furthermore, tracking the connections in each network between different colleges or even departments within a university can also help to examine the nature of interdisciplinary relations [195]. Thus, policy makers and administrators can be informed about the potentialities of the results found in this study, and they can interpret the results to formulate policies which will help to spur collaborative research across departmental and disciplinary boundaries. In the case of extending the study to the entire university, research performance can be determined based on collaboration relationships (e.g., density of networks or other structural properties of networks) between different sizes of universities (e.g., size can be specified according to the number of students, employees, departments, active facilities used in research, and etc.) to allocate research money to strengthen smaller universities that aspire to engage in collaborative research. If small-sized universities have just about the same relative amount of collaboration as large-sized universities (after capturing the in-progress collaborative relations via self-reported way), there will be no economies of scale in this matter [14].

Since this study aims at evaluating the extent to which social network metrics obtained from the researchers' multiple collaborative output networks as well as their communication networks predict the performance of researchers, the information obtained from this study can be used to formulate policies that improve both the collaborative and communication relationships that impact the performance of researchers. For example, when the level of prediction of eigenvector centrality on the performance of researchers is low, meaning that the researchers tend to both collaborate and communicate with other researchers that are not well connected (i.e., other researchers that are not well-performing in their collaborative activities and communications), policies could be generated, which primarily attempt to encourage the researchers to interact with other researchers who are active in their both collaborative and communication relationships.

By investigating the degree of the impact of researchers' individual innovativeness on their collaborative output, university administration will know the capability (i.e. the degree) of the different colleges, or even the university as a whole in case the study is extended to the entire university, in transforming the ideas embedded in researchers' networks into a productive work in a collaborative manner. Then, information concerning the extent to which researchers' individual innovativeness impacts their collaborative output can be used for the evaluation of different colleges in a university. In the case of low impact, university administration should initiate to devise policies, e.g., polices encouraging informal institutional arrangements, or programs in which informal group meetings occur to mediate the exchange of knowledge or ideas informally.

This study has three major limitations. First, the study intended to capture the in-progress collaborative relations in a self-reported way as well as the completed collaborative relations; however, there is an issue of accuracy when collecting self-reported data due to biased responses and poor memory [18, 44]. For example, respondents do not want to report collaborative output ties, especially joint patents, for confidentiality reasons. Moreover, it is highly possible that respondents might not remember all of their collaborative output ties, therefore they enter incomplete information. A future study can be made to compare the overlaps of the networks constructed by self-reported data with the networks constructed by database information. Despite these concerns, there are many recent studies using the self-report method [45-48]. Second, when this study is applied to other colleges and disciplines, some of these four networks disappear. For example, writing joint grant proposals in a college of business is not as common as in a college of engineering. Moreover, some colleges and disciplines such as college of education and business have a decreased tendency to issue patents, and in some disciplines such as humanities and history, single-authored papers are more valuable than co-authored papers. Furthermore, this study can be run for other colleges of engineering in different universities (e.g., small-sized or large-sized, research-oriented) to understand whether the findings of this study are more or less specific for the chosen sample. Third, selecting the values of base and α differently in the knowledge growth function, f(t), affects the output obtained from the function itself and the shape of the parabola capturing the growth rate of knowledge. Therefore, a sensitivity analysis can be run for the different values of KG which is obtained by using different f(t)s in order to understand how the results differ in the same model. Moreover, other types of f(t)s such as S-shaped functions can also be considered for knowledge growth.

REFERENCES

[1] H. F. Moed, W. Glänzel, and U. Schmoch, *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S & T Systems*. Dordrecht: Kluwer Academic Publishers, 2004.

[2] C. Freeman and L. Soete, "Developing science, technology and innovation indicators: What we can learn from the past," *Research Policy,* vol. 38, pp. 583-589, 2009.

[3] C. o. Competitiveness, "Innovate America: Thriving in a World of Challenge and Change," 2005.

[4] E. O. o. t. President, "A Strategy for American Innovation: Securing Our Economic Growth and Prosperity," 2011.

[5] C. o. Competitiveness, "Collaborate. Leading Regional Innovation Clusters," 2010.

[6] N. Hara, P. Solomon, S.-L. Kim, and D. H. Sonnenwald, "An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration," *Journal of the American Society for Information Science and Technology,* vol. 54, pp. 952-965, 2003.

[7] D. H. Sonnenwald, "Scientific collaboration," *Annual Review of Information Science and Technology,* vol. 41, pp. 643-681, 2007.

[8] J. S. Katz and B. R. Martin, "What is research collaboration?," *Research policy* vol. 26, pp. 1-18, 1997.

[9] G. Melin, "Pragmatism and self-organization: research collaboration on the individual level," *Research policy,* vol. 29, pp. 31-40, 2000.

[10] D. D. Beaver, "Reflections on scientific collaboration (and its study): past, present, and future," *Scientometrics,* vol. 52, pp. 365-377, 2001.

[11] H. Bukvova, "Studying Research Collaboration: A Literature Review," *Working Papers on Information Systems,* vol. 10, 2010.

[12] N. S. Board, "Research & Development, Innovation, and the Science and Engineering Workforce: A Companion to Science and Engineering Indicators 2012," National Science Foundation, Arlington, VA2012.

[13] K. Hale, "Collaboration in Academic R&D: A Decade of Growth in Pass-Through Funding. InfoBrief. NSF 12-325," National Science Foundation2012.

[14] G. Melin and O. Persson, "Studying research collaboration using co-authorships," *Scientometrics,* vol. 36, pp. 363-377, 1996.

[15] J. N. Cummings and S. Kiesler, "Collaborative research across disciplinary and organizational boundaries " *Social Studies of Science,* vol. 35, pp. 703-722, 2005.

[16] B. Bozeman and E. Corley, "Scientists' collaboration strategies: implications for scientific and technical human capital," *Research Policy,* vol. 33, pp. 599-616, 2004.

[17] M. F. Fox, "Publication Productivity among Scientists," *Social Studies of Science,* vol. 13, pp. 285-305, 1983.

[18] S. Lee and B. Bozeman, "The Impact of Research Collaboration on Scientific Productivity," *Social Studies of Science,* vol. 35, pp. 673-702, 2005.

[19] W. Glänzel and A. Schubert, "Analysing Scientific Networks through Co-authorship," in *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S & T Systems*, H. F. Moed, W. Glanzel, and U. Schmoch, Eds., ed Dordrecht: Kluwer Academic Publishers, 2004, pp. 257-276.

[20] J. Rigby, "Comparing the scientific quality achieved by funding instruments for single grant holders and for collaborative networks within a research system: Some observations," *Scientometrics,* vol. 78, pp. 145-164, 2009.

[21] M. Balconi, S. Breschi, and F. Lissoni, "Networks of inventors and the role of academia: an exploration of Italian patent data," *Research Policy,* vol. 33, pp. 127-145, 2004.

[22] S. Breschi and F. Lissoni, "Knowledge Networks from Patent Data," in *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S & T Systems*, H. F. Moed, W. Glanzel, and U. Schmoch, Eds., ed Dordrecht: Kluwer Academic Publishers, 2004, pp. 613-643.

[23] S. Breschi and F. Lissoni, "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows," *Journal of Economic Geography,* vol. 9, pp. 439-468, 2009.

[24] M. Meyer and S. Bhattacharya, "Commonalities and differences between scholarly and technical collaboration: An exploration of co-invention and co-authorship analyses," *Scientometrics,* vol. 61, pp. 443-456, 2004.

[25] M. E. Newman, "Scientific collaboration networks. I. Network construction and fundamental results," *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics,* vol. 64, pp. 016131-016131, 2001.

[26] M. E. Newman, "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality," *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics,* vol. 64, pp. 016132-016132, 2001.

[27] M. E. J. Newman, "The Structure of Scientific Collaboration Networks," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 98, pp. 404-409, 2001.

[28] A. L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations," *Physica A,* vol. 311, pp. 590-614, 2002.

[29] G. Laudel, "What do we measure by co-authorships?," *Research Evaluation,* vol. 11, pp. 3-15, 2002.

[30] W. O. Hagstrom, *The scientific community*. Carbondale: Southern Illinois University Press, 1975.

[31] T. D. Stokes and J. A. Hartley, "Coauthorship, Social Structure and Influence within Specialties," *Social Studies of Science,* vol. 19, pp. 101-125, 1989.

[32] M. C. LaFollette, *Stealing into print: fraud, plagiarism, and misconduct in scientific publishing*. Berkeley: University of California Press, 1992.

[33] A. Pepe, "The Relationship between Acquaintanceship and Coauthorship in Scientific Collaboration Networks," *Journal of the American Society for Information Science and Technology,* vol. 62, pp. 2121-2132, 2011.

[34] D. J. De Solla Price and D. D. Beaver, "Collaboration in an Invisible College," *American Psychologist,* vol. 21, pp. 1011-1018, 1966.

[35] D. Edge, "Quantitative Measures of Communication in Science: a Critical Review," *History of Science,* vol. 17, pp. 102-134, 1979.

[36] R. J. W. Tijssen, "Measuring and Evaluating Science-Technology Connections and Interactions: Towards International Statistics," in *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S & T Systems*, H. F. Moed, W. Glanzel, and U. Schmoch, Eds., ed Dordrecht: Kluwer Academic Publishers, 2004, pp. 695-715.

[37] R. Kraut and C. Egido, "Patterns of contact and communication in scientific research collaboration," in *ACM conference on Computer-supported cooperative work*, Portland, OR, 1988, pp. 1-12.

[38] G. M. Olson and J. S. Olson, "Distance Matters," *Human-Computer Interaction,* vol. 15, pp. 139-178, 2000.

[39] C. L. Borgman and J. Furner, "Scholarly Communication and Bibliometrics," *Annual Review of Information Science and Technology,* vol. 36, pp. 3-72, 2002.

[40] T. Schleyer, H. Spallek, B. S. Butler, S. Subramanian, D. Weiss, M. L. Poythress*, et al.*, "Facebook for scientists: Requirements and services for optimizing how scientific collaborations are established," *Journal of Medical Internet Research,* vol. 10, pp. 46-59, 2008.

[41] W. Glänzel, "Coauthorship patterns and trends in the sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies," *Library Trends*, vol. 50, pp. 461-473, 2002.

[42] H. Kretschmer, "Author productivity and geodesic distance in bibliographic co-authorship networks, and visibility on the Web," *Scientometrics,* vol. 60, pp. 409-420, 2004.

[43] H. Hou, H. Kretschmer, and L. I. U. Zeyuan, "The structure of scientific collaboration networks in Scientometrics," *Scientometrics,* vol. 75, pp. 189-202, 2008.

[44] E. Vasileiadou, "Stabilisation operationalised: Using time series analysis to understand the dynamics of research collaboration," *Journal of Informetrics,* vol. 3, pp. 36-48, 2009.

[45] R. B. Duque, M. Ynalvez, R. Sooryamoorthy, P. Mbatia, D.-B. S. Dzorgbo, and W. Shrum, "Collaboration paradox : Scientific productivity, the internet, and problems of research in developing areas," *Social Studies of Science*, vol. 35, pp. 755-785, 2005.

[46] R. Sooryamoorthy and W. Shrum, "Does the Internet Promote Collaboration and Productivity? Evidence from the Scientific Community in South Africa," *Journal of Computer-Mediated Communication,* vol. 12, pp. 733-751, 2007.

[47] F. J. Van Rijnsoever, L. K. Hessels, and R. L. J. Vandeberg, "A resource-based view on the interactions of university researchers," *Research Policy,* vol. 37, pp. 1255-1266, 2008.

[48] M. A. Ynalvez and W. M. Shrum, "Professional networks, scientific collaboration, and publication productivity in resource-constrained research institutions in a developing country," *Research Policy,* vol. 40, pp. 204-216, 2011.

[49] S. Wasserman and K. Faust, *Social network analysis: methods and applications*. Cambridge; New York: Cambridge University Press, 1994.

[50] J. E. Hirsch, "An Index to Quantify an Individual's Scientific Research Output," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 102, pp. 16569–16572, 2005.

[51] A. Abbasi, J. Altmann, and L. Hossain, "Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures," *Journal of Informetrics,* vol. 5, pp. 594-607, 2011.

[52] R. S. Burt, *Structural holes: the social structure of competition*. Cambridge, Mass. : Harvard University Press, 1992.

[53] J. Björk and M. Magnusson, "Where Do Good Innovation Ideas Come From? Exploring the Influence of Network Connectivity on Innovation Idea Quality," *Journal of Product Innovation Management,* vol. 26, pp. 662-670, 2009.

[54] P. B. Paulus and V. R. Brown, "Toward more creative and innovative group idea generation: A cognitive-social-motivational perspective of brainstorming," *Social and Personality Psychology Compass,* vol. 1, pp. 248-265, 2007.

[55] V. John-Steiner, *Creative collaboration* New York: Oxford University Press, 2000.

[56] W. S. Lovejoy and A. Sinha, "Efficient Structures for Innovative Social Networks," *Management Science,* vol. 56, pp. 1127-1145, 2010.

[57] H. T. Hurt, K. Joseph, and C. D. Cook, "Scales for the measurement of innovativeness," *Human Communication Research,* vol. 4, pp. 58-65, 1977.

[58] R. T. Keller and W. E. Holland, "Individual characteristics of innovativeness and communication in research and development organizations," *Journal of Applied Psychology,* vol. 63, p. 759, 1978.

[59] G. Cheney, B. L. Block, and B. S. Gordon, "Perceptions of innovativeness and communication about innovations: a study of three types of service organizations," *Communication Quarterly,* vol. 34, pp. 213-230, 1986.

[60] R. E. Goldsmith, "The validity of a scale to measure global innovativeness," *Journal of Applied Business Research (JABR),* vol. 7, pp. 89-97, 1991.

[61] R. F. Kleysen and C. T. Street, "Toward a multi-dimensional measure of individual innovative behavior," *Journal of intellectual Capital,* vol. 2, pp. 284-296, 2001.

[62] L. R. Flynn and R. E. Goldsmith, "A validation of the Goldsmith and Hofacker innovativeness scale," *Educational and Psychological Measurement,* vol. 53, pp. 1105-1116, 1993.

[63] R. McAdam, "Knowledge creation and idea generation: a critical quality perspective," *Technovation,* vol. 24, pp. 697-705, 2004.

[64] G. Szulanski, "Exploring Internal Stickiness: Impediments to the Transfer of Best Practice Within the Firm," *Strategic Management Journal,* vol. 17, pp. 27-43, Winter96 Special Issue 1996.

[65] B. Uzzi, "Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness," *Administrative Science Quarterly,* vol. 42, pp. 35-67, 1997.

[66] M. T. Hansen, "The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits," *Administrative Science Quarterly,* vol. 44, pp. 82-111, 1999.

[67] R. Reagans and B. McEvily, "Network Structure and Knowledge Transfer: The Effects of Cohesion and Range," *Administrative Science Quarterly,* vol. 48, pp. 240-267, 2003.

[68] K. A. Bollen and P. J. Curran, *Latent curve models : a structural equation perspective*. Hoboken, N.J.: John Wiley & Sons, 2006.

[69] B. Bozeman and J. D. Rogers, "A churn model of scientific knowledge value: Internet researchers as a knowledge value collective," *Research Policy,* vol. 31, pp. 769–794, 2002.

[70] M. S. Granovetter, "The strength of weak ties," *American Journal of Sociology,* vol. 78, pp. 1360-1380, 1973.

[71] P. V. Marsden and K. E. Campbell, "Measuring Tie Strength," *Social Forces,* vol. 63, pp. 482-501, 1984.

[72] J. Tague-Sutcliffe, "An introduction to informetrics," *Information Processing and Management* vol. 28, pp. 1-3, 1992.

[73] L. Egghe, "Expansion of the field of informetrics: Origins and consequences," *Information Processing and Management,* vol. 41, pp. 1311–1316, 2005.

[74] P. Ingwersen and L. Bjorneborn, "Methodological Issues of Webometric Studies," in *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems,* H. F. Moed, W. Glanzel, and U. Schmoch, Eds., ed: Dordrecht; Boston and London: Kluwer Academic, 2004, pp. 339-369.

[75] J. Bar-Ilan, "Informetrics at the beginning of the 21st century - A review," *Journal of Informetrics,* vol. 2, pp. 1-52, 2008.

[76] L. Björneborn and P. Ingwersen, "Toward a basic framework for webometrics," *Journal of the American Society for Information Science and Technology,* vol. 55, pp. 1216-1227, 2004.

[77] D. D. Beaver and R. Rosen, "Studies in scientific collaboration Part I. The professional origins of scientific co-authorship," *Scientometrics,* vol. 1, pp. 65-84, 1978.

[78] S. Wuchty, B. F. Jones, and B. Uzzi, "The increasing dominance of teams in production of knowledge," *Science (Wash.D.C.),* vol. 316, pp. 1036-1039, 2007.

[79] N. E. Friedkin, "University Social Structure and Social Networks among Scientists," *American Journal of Sociology,* vol. 83, pp. 1444-1465, 1978.

[80] H. P. F. Peters and A. F. J. Van Raan, "Structuring scientific activities by co-author analysis: an exercise on a university faculty level," *Scientometrics,* vol. 20, pp. 235-255, 1991.

[81] K. Subramanyam, "Bibliometric studies of research collaboration: A review," *Journal of Information Science,* vol. 6, pp. 33-38, 1983.

[82] T. A. Finholt, "Collaboratories as a New Form of Scientific Organization," *Economics of Innovation and New Technology,* vol. 12, pp. 5-25, 2003.

[83] R. R. Kraut, R. S. Fish, R. W. Root, and B. L. Chalfonte, "Informal communication in organizations: form, function, and technology," in *People's Reactions to Technology in Factories, Offices, and Aerospace,* O. S and S. S, Eds., ed Newbury Park, CA: Sage, 1990, pp. 145-199.

[84] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology,* vol. 27, pp. 415-444, 2001.

[85] P. V. Marsden, "Homogenity in confiding relations," *Social Networks,* vol. 10, pp. 57-76, 1988.

[86] B. F. Reskin, D. B. McBrier, and J. A. Kmec, "The determinants and consequences of workplace sex and race composition," *Annual Review of Sociology,* vol. 25, pp. 335-361, 1999.

[87] J. C. McCroskey. Available: http://www.jamescmccroskey.com/measures/innovation.htm

[88] D. A. Dillman, *Mail and internet surveys: the tailored design method*, 2nd ed. ed. Hoboken, N.J. : Wiley, 2007.

[89] S. P. Borgatti, M. G. Everett, and L. C. Freeman, "Ucinet for Windows: Software for Social Network Analysis," ed: Harvard, MA: Analytic Technologies., 2002.

[90] D. L. Hansen, B. Schneiderman, and M. A. Smith, *Analyzing Social Media Networks with NodeXL: Insights From a Connected World*. Burlington, MA: Morgan Kaufmann, 2011.

[91] R. C. Team, "R: A Language and Environment for Statistical Computing," ed. Vienna, Austria: R Foundation for Statistical Computing, 2012.

[92] M. Smith, N. Milic-Frayling, B. Shneiderman, E. Mendes Rodrigues, J. Leskovec, and C. Dunne, "NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010. http://nodexl.codeplex.com/ from the Social Media Research Foundation, http://www.smrfoundation.org.," ed, 2010.

[93] R. Hanneman and M. Riddle, *Introduction to social network methods*: University of California, Riverside, 2005.

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

[95] M. O. Jackson, *Social and economic networks*. Princeton, NJ: Princeton University Press, 2008.

[96] M. E. J. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, pp. 208701-208701, 2002.

[97] M. E. J. Newman, "Mixing patterns in networks," *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 67, pp. 026126-026126, 2003.

[98] S. Borgatti, "Identifying sets of key players in a social network," *Computational & Mathematical Organization Theory*, vol. 12, p. 21, 2006.

[99] T. A. B. Snijders and S. P. Borgatti, "Non-parametric standard Errors and Tests for Network Statistics " *Connections*, vol. 22, pp. 61-70, 1999.

[100] D. Krackhardt, "QAP partialling as a test of spuriousness," *Social Networks*, vol. 9, pp. 171-186, 1987.

[101] N. Mantel, "The detection of disease clustering and a generalized regression approach," *Cancer Research*, vol. 27, pp. 209–220, 1967.

[102] L. J. Hubert, *Assignment methods in combinatorial data analysis*. New York: M. Dekker, 1987.

[103] D. Dekker, D. Krackhardt, and T. A. B. Snijders, "Sensitivity of MRQAP tests to Colllinearity and Autocorrelation conditions," *Psychometrika*, vol. 72, pp. 563-581, 2007.

[104] L. A. Waller and C. A. Gotway, *Applied Spatial Statistics for Public Health Data*. Hoboken, N.J.: John Wiley & Sons, 2004.

[105] D. Krackhardt, "Predicting with networks: Nonparametric multiple regression analysis of dyadic data," *Social Networks,* vol. 10, pp. 359-381, 1988.

[106] M. Kilduff and D. Krackhardt, "Bringing the individual back in: A structural analysis of the internal market for reputation in organizations," *Academy of Management Journal,* vol. 37, pp. 87-108, 1994.

[107] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, "An introduction to exponential random graph (p*) models for social networks," *Social Networks,* vol. 29, pp. 173-191, 2007.

[108] G. Robins, T. Snijder, W. Peng, M. Handcock, and P. Pattison, "Recent developments in exponential random graph (p*) models for social networks: Advances in exponential random graph (p*) models," *Social Networks,* vol. 29, pp. 192-215, 2007.

[109] S. M. Goodreau, M. S. Handcock, D. R. Hunter, C. T. Butts, and M. Morris, "A statnet Tutorial," *Journal of Statistical Software,* vol. 24, pp. 1-27, 2008.

[110] D. R. Hunter, M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris, "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks," *Journal of Statistical Software,* vol. 24, pp. 1-29, 2008.

[111] M. Morris, M. S. Handcock, and D. R. Hunter, "Specifcation of Exponential-Family Random Graph Models: Terms and Computational Aspects," *Journal of Statistical Software,* vol. 24, pp. 1-24, 2008.

[112] T. A. B. Snijders, "Statistical Models for Social Networks," *Annual review of sociology,* vol. 37, pp. 131-153, 2011.

[113] S. Wasserman and P. Pattison, "Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs and p*," *Psychometrika,* vol. 61, pp. 401-25, 1996.

[114] R. B. Freeman and W. Huang, "Collaborating With People Like Me: Ethnic co-authorship within the US," National Bureau of Economic Research No. 19905, 2014.

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

[116] S. P. Borgatti, "Centrality and Network Flow " *Social Networks,* vol. 27, pp. 55-71, 2005.

[117] P. Bonacich, "Factoring and Weighting Approaches to Status Scores and Clique Identification," *Journal of Mathematical Sociology,* vol. 2, pp. 113-120, 1972.

[118] S. P. Borgatti and E. M. G., "Network analysis of 2-mode data," *Social Networks,* vol. 19, pp. 243-269, 1997.

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

[120] S. P. Borgatti. Available: http://www.analytictech.com/mb119/chapter5.htm

[121] M. G. Everett and S. P. Borgatti, "The centrality of groups and classes," *Journal of Mathematical Sociology*, vol. 23, pp. 181-201, 1999.

[122] Y. Jiang, "Locating active actors in the scientific collaboration communities based on interaction topology analyses," *Scientometrics*, vol. 74, pp. 471-485, 2008.

[123] D. Defazio, A. Lockett, M. Wright, L. Bornmann, and H. D. Daniel, "Funding Incentives, Collaborative Dynamics and Scientific Productivity: Evidence from the EU Framework Program," *Research Policy*, vol. 38, pp. 293-305, 2009.

[124] L. Bornmann and H. D. Daniel, "The state of h index research Is the h index the ideal way to measure research performance?," EMBO reports 1469221X, 2009.

[125] L. Bornmann and H.-D. Daniel, "What do we know about the h index?," *Journal of the American Society for Information Science and Technology*, vol. 58, pp. 1381-1385, 2007.

[126] R. Costas and M. Bordons, "The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level," *Journal of Informetrics*, vol. 1, pp. 193-203, 2007.

[127] B. Cronin and L. Meho, "Using the h-index to rank influential information scientists," *Journal of the American Society for Information Science and Technology*, vol. 57, pp. 1275-1278, 2006.

[128] J. E. Hirsch, "Does the h-index have predictive power?," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 19193–19198, 2007.

[129] L. Bornmann, R. Mutz, and H.-D. Daniel, "Are There Better Indices for Evaluation Purposes than the h Index? A Comparison of Nine Different Variants of the h Index Using Data from Biomedicine," *Journal of the American Society for Information Science and Technology*, vol. 59, pp. 830-837, 2008.

[130] C. McCarty, J. W. Jawitz, A. Hopkins, and A. Goldman, "Predicting author h-index using characteristics of the co-author network," *Scientometrics*, vol. 96, pp. 467-483, 2013.

[131] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 7821-7826, 2002.

[132] D. W. Aksnes, "Characteristics of highly cited papers," *Research Evaluation*, vol. 12, pp. 159-170, 2003.

[133] S. P. Borgatti, "Structural Holes: Unpacking Burt's Redundancy Measures," *Connections*, vol. 20, pp. 35-38, 1997.

[134] T. T. Baldwin, M. D. Bedell, and J. L. Johnson, "The social fabric of a team-based MBA Program: Network effects on student satisfaction and performance," *Academy of Management Journal*, vol. 40, pp. 1369-1397, 1997.

[135] A. Mehra, M. Kilduff, and D. J. Brass, "The Social Networks of High and Low Self-Monitors: Implications for Workplace Performance," *Administrative Science Quarterly*, vol. 46, pp. 121-146, 2001.

[136] R. T. Sparrowe, R. C. Liden, S. J. Wayne, and M. L. Kraimer, "Social Networks and the Performance of Individuals and Groups," *Academy of Management Journal*, vol. 44, pp. 316-325, 2001.

[137] A. Mehra, A. L. Dixon, D. J. Brass, and B. Robertson, "The Social Network Ties of Group Leaders: Implications for Group Performance and Leader Reputation," *Organization Science*, vol. 17, pp. 64-79, 2006.

[138] J. M. Hilbe, *Negative Binominal Regression* vol. Second Edition. New York: Cambridge University Press, 2011.

[139] A. C. Cameron and P. K. Trivedi, *Regression analysis of count data*. Cambridge, UK: Cambridge University Press,, 1998.

[140] G. Rodríguez. (2007). *Lecture Notes on Generalized Linear Models*. Available: http://data.princeton.edu/wws509/notes/

[141] B. G. Tabachnick and L. S. Fidell, *Using multivariate statistics* 5th ed. ed. Boston Pearson/Allyn & Bacon, 2007.

[142] (November 24). *Introduction to SAS*. Available: http://www.ats.ucla.edu/stat/sas/notes2/

[143] N. A. Rashwan and M. M. Kamel, "Using generalized Poisson log linear regression models in analyzing two-way contingency tables," *Applied Mathematical Sciences*, vol. 5, pp. 213-222, 2011.

[144] A. Baccini, L. Barabesi, M. Marcheselli, and L. Pratelli, "Statistical inference on the h-index with an application to top-scientist performance," *Journal of Informetrics*, vol. 6, pp. 721-728, 2012.

[145] P. B. Paulus, "Groups, Teams, and Creativity: The Creative Potential of Idea-generating Groups," *Applied Psychology: An International Review*, vol. 49, pp. 237-262, 2000.

[146] T. L. Albrecht and B. Hall, "Relational and Content Differences Between Elites and Outsiders in Innovation Networks," *Human Communication Research,* vol. 17, pp. 535-561, 1991.

[147] M. W. H. Weenig, "Communication Networks in the Diffusion of an Innovation in an Organization1," *Journal of Applied Social Psychology,* vol. 29, pp. 1072-1092, 1999.

[148] J. Kratzer, O. T. A. J. Leenders, and J. M. L. v. Engelen, "Stimulating the Potential: Creative Performance and Communication in Innovation Teams," *Creativity and Innovation Management,* vol. 13, pp. 63-71, 2004.

[149] K. G. Smith, C. J. Collins, and K. D. Clark, "Existing Knowledge, Knowledge Creation Capability, and the Rate of New Product Introduction in High-Technology Firms," *The Academy of Management Journal*, vol. 48, pp. 346-357, 2005.

[150] E. M. Rogers, *Diffusion of innovations,* 4th ed. ed. New York: Free Press, 1995.

[151] P. B. Paulus and B. A. Nijstad, *Group creativity: Innovation through collaboration*. New York, NY: Oxford University Press, 2003.

[152] J. S. Coleman, "Social capital in the creation of human capital," *The American Journal of Sociology,* vol. 94, pp. 95-120, 1988.

[153] M. Gargiulo and M. Benassi, "Trapped in Your Own Net? Network Cohesion, Structural Holes, and the Adaptation of Social Capital," *Organization Science,* vol. 11, pp. 183-196, 2000.

[154] K. Rost, "The strength of strong ties in the creation of innovation," *Research Policy,* vol. 40, pp. 588-604, 2011.

[155] N. Lin, "Building a network theory of social capital," *Connections*, vol. 22, pp. 28-51, 1999.

[156] R. S. Burt, "Structural holes versus network closure as social capital," in *Social capital: Theory and research*, N. Lin, K. S. Cook, and R. S. Burt, Eds., ed New Brunswick, New York: Transaction Publishers, 2001, pp. 31-56.

[157] P. S. Adler and S.-W. Kwon, "Social Capital: Prospects for a New Concept," *The Academy of Management Review,* vol. 27, pp. 17-40, 2002.

[158] G. Ahuja, "Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study," *Administrative Science Quarterly,* vol. 45, pp. 425-455, 2000.

[159] R. S. Burt, "Structural Holes and Good Ideas," *American Journal of Sociology,* vol. 110, pp. 349-399, 2004.

[160] R. Cowan and N. Jonard, "Structural holes, innovation and the distribution of ideas," *Journal of Economic Interaction and Coordination,* vol. 2, pp. 93-110, 2007.

[161] R. Cowan and N. Jonard, "Network structure and the diffusion of knowledge," *Journal of economic dynamics and control,* vol. 28, pp. 1557-1575, 2004.

[162] J. Hemphälä and M. Magnusson, "Networks for Innovation–But What Networks and What Innovation?," *Creativity and Innovation Management,* vol. 21, pp. 3-16, 2012.

[163] M. Du Plessis, "The role of knowledge management in innovation," *Journal of knowledge management,* vol. 11, pp. 20-29, 2007.

[164] M. Alavi and D. E. Leidner, "Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues," *MIS Quarterly,* vol. 25, pp. 107-136, 2001.

[165] D. F. Midgley and G. R. Dowling, "Innovativeness: The Concept and Its Measurement," *Journal of Consumer Research,* vol. 4, pp. 229-242, 1978.

[166] D. Z. Levin and R. Cross, "The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer," *Management science,* vol. 50, pp. 1477-1490, 2004.

[167] N. E. Friedkin, "Information flow through strong and weak ties in intraorganizational social networks," *Social networks,* vol. 3, pp. 273-285, 1982.

[168] G. Weimann, "The strength of weak conversational ties in the flow of information and influence," *Social Networks,* vol. 5, pp. 245-267, 1983.

[169] D. Krackhardt, "The strength of strong ties: The importance of philos in organizations," in *Networks and organizations: Structure, form, and action,* ed, 1992, pp. 216-239.

[170] M. Ruef, "Strong ties, weak ties and islands: structural and cultural predictors of organizational innovation," *Industrial and Corporate Change,* vol. 11, pp. 427-449, June 1, 2002 2002.

[171] K. M. Mathews, M. C. White, R. G. Long, B. Soper, and C. W. Von Bergen, "Association of indicators and predictors of tie strength," *Psychological Reports,* vol. 83, pp. 1459-1469, 1998.

[172] A. Petróczi, T. Nepusz, and F. Bazsó, "Measuring tie-strength in virtual social networks," *Connections,* vol. 27, pp. 39-52, 2007.

[173] C. M. Ringle, S. Wende, and A. Will, "SmartPLS (beta)," 2.0.M3 ed. Hamburg, Germany, 2005.

[174] M. Haenlein and A. M. Kaplan, "A Beginner's Guide to Partial Least Squares Analysis," *Understanding Statistics,* vol. 3, pp. 283-297, 2004.

[175] N. Urbach and F. Ahlemann, "Structural Equation Modeling in Information Systems Research Using Partial Least Squares," *Journal of Information Technology Theory and Application,* vol. 11, pp. 5-40, 2010.

[176] A. Monecke and F. Leisch, "semPLS: Structural Equation Modeling Using Partial Least Squares," *Journal Of Statistical Software,* vol. 48, pp. 1-32, 2012.

[177] H. Wold, "Partial Least Squares," in *Encyclopedia of Statistical Sciences*. vol. 6, N. J. S Kotz, Ed., ed New York: John Wiley & Sons, 1985, pp. 581-591.

[178] K. Jöreskog, "Structural analysis of covariance and correlation matrices," *Psychometrika,* vol. 43, pp. 443-477, 1978.

[179] M. Tenenhaus, V. Esposito Vinzi, Y.-M. Chatelin, and C. Lauro, "PLS path modeling," *Computational Statistics & Data Analysis,* vol. 48, pp. 159-205, 2005.

[180] W. Chin and P. R. Newsted, "Structural Equation Modeling analysis with Small Samples Using Partial Least Squares," in *Statistical Strategies for Small Sample Research*, R. Hoyle, Ed., ed Thousand Oaks, CA: Sage Publications, 1999, pp. 307-341.

[181] M. Wetzels, G. Odekerken-Schröder, and C. Van Oppen, "Using PLS Path Modeling for Assessing Hierarchical Construct Models: Guidelines and Empirical Illustration " *MIS Quarterly,* vol. 33, pp. 177-195, 2009.

[182] J. Henseler, C. M. Ringle, and R. R. Sinkovics, "The use of partial least squares path modeling in international marketing " *Advances in International Marketing,* vol. 8, pp. 277-319, 2009.

[183] W. Chin, "How toWrite Up and Report PLS Analyses," in *Handbook of partial least squares: concepts, methods and applications* V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang, Eds., ed Berlin ; New York: Springer, 2010, pp. 655-690.

[184] V. Esposito Vinzi, L. Trinchera, and S. Amato, "PLS Path Modeling: From Foundations to Recent Developments and Open Issues for Model Assessment and Improvement," in *Handbook of partial least squares: concepts, methods and applications* V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang, Eds., ed Berlin; New York: Springer, 2010, pp. 47-82.

[185] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika,* vol. 16, pp. 297-334, 1951.

[186] W. Chin, "The partial least squares approach to structural equation modeling," in *Modern methods for business research*, G. A. Marcoulides, Ed., ed Mahwah, NJ: LawrenceErlbaumAssociates, 1998, pp. 295-336.

[187] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. NewYork,NY: Chapman Hall, 1993.

[188] C. Fornell and D. F. Larcker, "Structural equation models with unobservable variables and measurement error: Algebra and statistics," *Journal of Marketing Research,* vol. 18, pp. 382-388, 1981.

[189] J. F. Hair, C. M. Ringle, and M. Sarstedt, "PLS-SEM: Indeed a silver bullet," *The Journal of Marketing Theory and Practice,* vol. 19, pp. 139-152, 2011.

[190] E. A. Smith, "The role of tacit and explicit knowledge in the workplace," *Journal of knowledge Management,* vol. 5, pp. 311-321, 2001.

[191] I. Nonaka, "A dynamic theory of organizational knowledge creation," *Organization science,* vol. 5, pp. 14-37, 1994.

[192] I. Nonaka and G. Von Krogh, "Perspective-tacit knowledge and knowledge conversion: Controversy and advancement in organizational knowledge creation theory," *Organization science,* vol. 20, pp. 635-652, 2009.

[193] R. T. Herschel, H. Nemati, and D. Steiger, "Tacit to explicit knowledge conversion: knowledge exchange protocols," *Journal of knowledge management,* vol. 5, pp. 107-116, 2001.

[194] J. Henseler and G. Fassott, "Testing Moderating Effects in PLS Path Models: An Illustration of Available Procedures," in *Handbook of Partial Least Squares*, V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang, Eds., ed Berlin; New York: Springer, 2010, pp. 713-735.

[195] C. Sá, "‘Interdisciplinary strategies’ in U.S. research universities," *Higher Education,* vol. 55, pp. 537-552, 2008.

APPENDICES

Appendix A1: A Questionnaire to Collect the Researchers' Collaborative Output Ties (First Page)

FIRST NAME: LAST NAME: COUNTRY of ORIGIN:

STEP 1 (COLLABORATION INFORMATION)

Q1) With whom do you collaborate for your research matters? And

Q2) How many completed and uncompleted collaborative work do you have with other researchers including

- in-preparation, (re)submitted or rejected, and published **joint publications (column 1)?** **(1)** for 1-2, **(2)** for 3-5, **(3)** for 6-9, **(4)** for 10-above
- in-preparation, declined, and funded **grant proposals (column 2)?** **(1)** for 1-2, **(2)** for 3-5, **(3)** for 6-9, **(4)** for 10-above
- including rejected, submitted, and issued **patent applications (column 3)?** **(1)** for 1-2, **(2)** for 3-5, **(3)** for 6-9, **(4)** for 10-above

Chemical & Biomedical Engineering

Last Name	Name	Public.	Grant	Patent

Civil & Environmental Engineering

Last Name	Name	Public.	Grant	Patent

Computer Science & Engineering

Last Name	Name	Public.	Grant	Patent

Electrical Engineering

Last Name	Name	Public.	Grant	Patent

Industrial & Management Systems Engineering

Last Name	Name	Public.	Grant	Patent

Mechanical Engineering

Last Name	Name	Public.	Grant	Patent

College of Engineering Dean

Last Name	Name	Public.	Grant	Patent

Please also write a name from other USF colleges or institutions below:

Last Name	Name	Department	Public.	Grant	Patent

Appendix A2: A Questionnaire to Collect the Researchers' Communication Ties (Second Page)

STEP 2 (COMMUNICATION INFORMATION)

Q1) With whom do you exchange conversations or ideas via below mentioned ways?

Face-to-Face Conversations:

1) formal or informal group meetings and events in DEPARTMENT, COLLEGE, and even CAMPUS level
2) hallway conversations in DEPARTMENT and COLLEGE level
3) serving in a student's doctoral committee
4) telephone conversations, and etc.

Conversations in Virtual Environment:

1) e-mail exchange
2) exchanging ideas in online social network (academia.edu), and etc.

Q2) How frequently do you exchange conversations or ideas?

once a day**(6),** once a week**(5),** once every two week**(4),** once a month**(3),** once every two months**(2),** once every three months**(1)**

Q3) How close is your relationship between you and your conversational partner?

Very Close**(6),** Close**(5),** Somewhat close**(4),** Somewhat Distant**(3),** Distant**(2),** Very Distant**(1)**

Q4) How often do you discuss your work or home personal problems with your conversational partner?

Very Often **(5)**, Often **(4)**, Occasionally **(3)**, Seldom **(2)**, Never **(1)**

Chemical & Biomedical Engineering

Last Name	Name	Q2	Q3	Q4

Civil & Environmental Engineering

Last Name	Name	Q2	Q3	Q4

Computer Science & Engineering

Last Name	Name	Q2	Q3	Q4

Electrical Engineering

Last Name	Name	Q2	Q3	Q4

Industrial & Management Systems Engineering

Last Name	Name	Q2	Q3	Q4

Mechanical Engineering

Last Name	Name	Q2	Q3	Q4

College of Engineering Dean

Last Name	Name	Q2	Q3	Q4

Please also write a name from other USF colleges or institutions below:

Last Name	Name	Department	Q2	Q3	Q4

Appendix A3: A Questionnaire to Measure Researchers' Self-Perceived Innovativeness (Third Page)

Please indicate the degree to which each statement applies to you by marking whether you: **Strongly Disagree (1); Disagree (2); Neutral (3); Agree (4); Strongly Agree (5).**

1. ____ My peers often ask me for advice or information
2. ____ I enjoy trying new ideas.
3. ____ I seek out new ways to do things.
4. ____ I am generally cautious about accepting new ideas.
5. ____ I frequently improvise methods for solving a problem when an answer is not apparent.
6. ____ I am suspicious of new inventions and new ways of thinking.
7. ____ I rarely trust new ideas until I can see whether the vast majority of people around me accept them.
8. ____ I feel that I am an influential member of my peer group.
9. ____ I consider myself to be creative and original in my thinking and behavior.
10. ____ I am aware that I am usually one of the last people in my group to accept something new.
11. ____ I am an inventive kind of person.
12. ____ I enjoy taking part in the leadership responsibilities of the group I belong to.
13. ____ I am reluctant about adopting new ways of doing things until I see them working for people around me.
14. ____ I find it stimulating to be original in my thinking and behavior.
15. ____ I tend to feel that the old way of living and doing things is the best way.
16. ____ I am challenged by ambiguities and unsolved problems.
17. ____ I must see other people using new innovations before I will consider them.
18. ____ I am receptive to new ideas.
19. ____ I am challenged by unanswered questions.
20. ____ I often find myself skeptical of new ideas.

Appendix B: Image of the Copyright Permission for the Third Page of Appendix A

This is a License Agreement between Oguz Cimenler ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3376130027816
License date	Apr 25, 2014
Licensed content publisher	John Wiley and Sons
Licensed content publication	Human Communication Research
Licensed content title	SCALES FOR THE MEASUREMENT OF INNOVATIVENESS
Licensed copyright line	Copyright © 2006, John Wiley and Sons
Licensed content author	H. THOMAS HURT,KATHERINE JOSEPH,CHESTER D. COOK
Licensed content date	Mar 17, 2006
Start page	58
End page	65
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print
Portion	Text extract
Number of Pages	1
Will you be translating?	No
Title of your thesis / dissertation	Social Network Analysis of Researchers' Communication and Collaborative Networks Using Their Self-reported Data
Expected completion date	Aug 2014
Expected size (number of pages)	142
Total	0.00 USD

Appendix C: Image of the Written Permission for Published Portion of Chapter 3

This is a License Agreement between Oguz Cimenler ("You") and Elsevier ("Elsevier"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

Get the printable license.

License Number	3417590425411
License date	Jun 28, 2014
Licensed content publisher	Elsevier
Licensed content publication	Journal of Informetrics
Licensed content title	A regression analysis of researchers' social network metrics on their citation performance in a college of engineering
Licensed content author	Oguz Cimenler,Kingsley A. Reeves,John Skvoretz
Licensed content date	July 2014
Licensed content volume number	8
Licensed content issue number	3
Number of pages	16
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	Social Network Analysis of Researchers' Communication and Collaborative Networks Using Their Self-reported Data
Expected completion date	Aug 2014
Estimated size (number of pages)	142
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD

ABOUT THE AUTHOR

Oguz Cimenler received his B.S. degree in Textile Engineering from University of Gaziantep, Turkey in 2003, and his M.B.A degree from Auburn University-Montgomery, Alabama in 2006. Since 2009, he has been working on his Ph.D. in Industrial & management Systems Engineering at the University of South Florida (USF). Until 2008, he worked as a Product Development Coordinator & Cost Analyst for VF Corporation located in Tampa, FL, which is one of the largest apparel producers in the world. Furthermore, he has also worked as data analyst for the Center of Transformation and Innovation at USF Research Park between 2012 and 2013. His main interests are analytics, multivariate statistics and social network analysis.