TOWSON UNIVERSITY

OFFICE OF GRADUATE STUDIES

Exploring Individual Status and Collaboration: Three Studies using Social Network Analysis Methods

By

Jameson McFarlane

A Dissertation

Presented to the faculty of Towson University

in partial fulfillment of the requirements for the degree

Doctor of Science

Department of Computer and Informational Sciences

Towson University

Towson, Maryland 21252

December, 2015

TOWSON UNIVERSITY

COLLEGE OF GRADUATE STUDIES AND RESEARCH

DISSERTATION APPROVAL PAGE

This is to certify that the thesis prepared by Jameson McFarlane entitled "Exploring Individual Status and Collaboration: Three Studies using Social Network Analysis Methods" has been approved by this committee as satisfactory completion of the requirement for the degree of Doctor of Science in Information Technology

Dr. Siddharth Kaza Chair, Thesis Committee	Signature	12 2 2015 Date
Dr. Heidi Feng Committee Member	Signature	12/02/2015 Date
Dr. Yanggon Kim Committee Member	Signature	12-02-2015 Date
Dr. Josh Dehlinger Committee Member	Signature	12-2-2015 Date
Dr. Mike McGuire Committee Member	Signature	12-02-2015 Date
Dr. Ramesh K. Karne Doctoral Program Director	Signature	12-02-15 Date
Dr. Chao Lu Chair, CIS	Signature	12/10/2015 Date
Dr. Janet DeLany Dean, College of Graduate Studies and Research	Signature	1-5-16 Date

Acknowledgements

Indeed, a dissertation is a significant benchmark in anyone's academic career. Mine is no exception. This tremendous accomplishment would not have been realized were it not through the artful leadership, and amiable support and inspiration of some incredible people at each stage of this research process.

I would like to extend my sincere gratitude to Dr. Kaza, my advisor, who served as the chair of my dissertation committee for his incredible patience, his caring attitude, as well his distinctive ability to see and understand the big picture of research development and analysis. Moreover, I would also like to thank Dr. Kaza for his painstaking guidance and sincere recommendations for completing the dissertation successfully.

In addition, I would also like to thank all the members of the dissertation committee: Dr. KimYanggon; Dr. Josh Dehlinger; Dr. Feng Jinjuan and Dr. Michael McGuire for their guidance, their recommendations and their ability to communicate the rationale for each step of the dissertation phase.

I would remiss if I don't express my most sincere gratitude to my wife, Anna; my sons, Thorne and Khalil; and my siblings for their kind support, heartfelt encouragement and sincere devotions during this particularly challenging time of my life. I know my parents are looking down upon me throughout each chapter in the research. Thank you and keep watching over me.

To all, I am eternally grateful.

ABSTRACT

Among the most trendy and ubiquitous services provided on the web are online social networking sites such as Twitter, Facebook, and LinkedIn to name a few. Participants of these social networking sites form and engage in a complex network system where individuals share information about their daily activities; keep abreast with families, friends, and acquaintances; as well as share and distribute knowledge and informational content to both their ‘friends’ in the network. The dynamic nature of online social networks and the dramatic increase in their popularity since the last decade provide a large-scale data store for the study of the structure, patterns of behavior, relationship roles and other resultant characteristics of the network graph among the social entities within the social network.

This dissertation that examines various aspects of Social Network Analysis (SNA) contains one static SNA essay and two dynamic SNA cases studies. In the first essay, the focus is on the structure of online social networks, specifically Twitter, and how sitting United States governors utilize the social network to distribute information to citizens. Many of the key conceptual theories of static SNA including structural balance, transitivity, reciprocity, social cohesion, influence, dominance, conformity and social role are underscored. The study provides evidence that although transitivity and reciprocity occur with a high level of interaction, there is very little dominance in the structure of the network. The result supports other studies in this area. In addition, the study provides further evidence to support structural balance (or imbalance) of geographical homophily since the majority of the friends and followers of these governors are from their state. Moreover, the results indicate a network imbalance resulting in the isolation of the dissemination of resources and services to citizens.

In the first case study, SNA graph theory is used to analyze the evolution and collaborations of cybersecurity education researchers in the domain of academic research. The primary emphasis is on analytic principles and concepts and the use of graph theory to represent the network data. Further, using the graph theories, we utilized SNA analytic tools to help make predictions about the principal network structure. This case study discusses many of the primary concepts of interpreting patterns of social ties among individuals in the network community. We examined the patterns of interactions among members of the network group to determine the structure of the network as well as the existence of cohesive sub-groups within the main network community. In addition, we analyzed the network to ascertain which central figure(s) play(s) key roles in the social network community. Not only is the structural prestige of a person in the social network a clear indication of the stability of the network but also a signal of the importance of social ties in the diffusion of information throughout the social system. Finally, we examined the social network to observe underlying correlating factors that have influenced the structure of the network with time.

The second case study also uses SNA graph theory to analyze the evolution of research collaborations of Information Assurance Education (IAE), Security Education (SE) and Cybersecurity Education (CSE) researchers in primarily an academic domain. One of the primary objectives of this case study is to use a dynamic approach to explain the static topological features of the researchers in the collaborative network using a time framework. Within the past few decades, scholars and experts from the government, academia, and businesses from various local, regional and international institutions have collaborated on research topics in IAE, SE

and CSE. The goal of this study is to analyze the research collaborations of IAE (that includes SE and CSE) educators who have published work in IEEE and ACM publications and other venues indexed by the IEEE and ACM digital libraries between 1999 and 2013. Specifically, we examine the structure of co-authorship in the IEEE and ACM community using various social network analysis (SNA) techniques and SNA tools. Overall, our results revealed a weakly connected IEEE and ACM network. Further, we saw a moderate increase in research collaboration at both institutional and cross-institutional levels. The results also show that the most prolific institutions are educational institutions related with the military.

Table of Contents

List of Figures

List of Tables

CHAPTER 1: INTRODUCTION

Networks are made up of objects, whether people or things, that connect with each other in various patterns and structures. In real life, we all have family members, circle of friends, or perhaps church members that make our social network. In the real world, we have observed a proliferation of schemes taking the form of social networks. This includes the World Wide Web, biological networks, food webs, information networks, technological networks, to name a few [1]. For decades, there have been a significant and emergent public and academic interest in social networks which can be attributed to the release of the acclaimed film and play *"Six Degree of Separation"* as well as the popular web game *"The Oracle of Bacon"* [2]. With the growing popularity of online social networking sites, such as Twitter and Facebook, as well as the ensuing amount of readily available data has generated a large scale interest to study the characteristics and structure of these social networks using various social network graphs and other analytic methods; an approach commonly referred to as social network analysis (SNA).

1.1 What is Social Network Analysis?

Social Network Analysis (SNA) is the study of the structure of social relations among relationships in a group to uncover the patterns and implications of these relationships [3]–[5]. Social network analysis offers the methodology to conceptualize social networks and analyze them with the ultimate objective of detecting and interpreting the patterns of ties among entities, or actors [5].

1.2 Background of social network analysis

The majority of researchers have attributed the contribution of Joseph Moreno to the field of *sociometry* – the measurement of interpersonal relations in small groups - in the 1930s as the precursor to SNA [3]. Moreno's sociometry approach utilizes *sociograms* which are visual depictions of any social entity and the relationships linking those individuals [4]. The acknowledgment that visual displays including sociograms can be used to study social structure advanced the development of rapid analytic techniques and methodologies to represent social network data and mathematical approaches to study social systems [3].

In the 1950s, an MIT researcher named Alex Bavelas, experimented on a small group of actors to determine how information travelled within the group as well as which network structures affected the speed and efficiency of information diffusion [4], [6]. This experiment led to the very significant concept for network analysis called *centrality*. The findings also led to the other important concepts of social structure: *reciprocity* and *mutuality*. In addition to the aforementioned concepts, numerous other concepts were postulated including the graph entity of *clique*; *structural balance*; *transitivity*; *social status*; *social role*; *structural equivalence*; and *social position*.

1.3 The network perspective

The most differentiating characteristics of the social network perspective is that social network analysis focuses on relationships among social entities and on the patterns and implications of these relationships [3] . Hence, rather than analyzing individual behaviors, attitudes and beliefs, SNA's focal point is on social entities or actors as they interact with each other and how those interaction constitute a framework that can be studied and analyzed.

Moreover, the social network perspective makes a variety of assumptions about actors, relations, and the resulting structure [3]:

- Actors and their actions are viewed as interdependent rather than independent, autonomous units.
- Relational ties (linkages) between actors are channels for transfer or “flow” of resources (either material, like money, or nonmaterial, like information, political support, friendship, or respect).
- Network models focusing on individuals view the network structural environment as providing opportunities for or constraints on individual action.
- Network models conceptualize structure (whether social, economic, political, and so forth) as enduring patterns of relations among actors.

Therefore, the fundamental theory behind social network analysis is to examine relational ties: whether it constitute *dyads* (the most basic level, establishing a link between two individuals); *triads* (a subset of three actors); or *subgroups* (any subset of actors and all the ties among them).

1.4 Modern theories

Whereas the bulk of the work on social network analysis was being fashioned out in the 1960s by Harvard students [4], lately there have been much theoretical work on the properties of complex graphs. These include the *small-world* theory, the *exponential random graph model* (ERGM), *power-law*, and *scale-free* networks.

In the 1960s, research on the *Small World* problem, in particular by the well-known social psychologist Stanley Milgram, provided empirical evidence on the ties of acquaintances - suggesting that almost every human on earth is separated by "six degrees of separation" [7]. Proponents of the small-world phenomenon argue that large networks have a small diameter and exhibit high clustering. More recently studies have revealed that the web, social networks, and even scientific research papers, exhibit some form of the small-world characteristics [8] .

There have been a substantial amount of research on random graphs beginning with the study by Erdös and Réyni [9], [10]. More recent work on using statistical models for analyzing social network data has led to a family of models passionately known as the *exponential random graph models* (ERGM) which includes *p1*, *p2*, and *p**. With ERGM, the analyst assumes existing dependencies among ties in network nodes. Of all the class of ERGM models, p* is the most commonly used [11], [12]. In addition, it is the preferred model for making statistical inferences on cross-sectional network data [4].

In very large and complex networks, the distribution may have more sample data with extreme values than normal distributions, drawing a curve with a long tail lowering as the value increases. Such distribution characteristic is called a *power-law*. Research has found that many real-world networks including the Internet topologies [13], the world-wide web [14], and to a significant extent social networks [15], have a power-law structure in nature.

A scale-free network is a special category of a power-law network in which the high-degree nodes are connected to other high-degree nodes [16]. Scale-free networks are extremely heterogeneous and their topology is dominated by a few highly connected nodes which link the rest of the less connected nodes to the system [17].

1.5 Models in SNA

There is a myriad of techniques for measuring the properties of social networks. But why use statistical models? Robins [18] highlighted the following reasons for using statistical models:

- Since social behavior is complex, stochastic models allow us to capture both the regularities in the processes giving rise to network ties while at the same time recognizing that there is variability that we are unlikely to be able to model in details. Most importantly, a well-specified stochastic model allows us to understand the uncertainty associated with observed outcomes.
- Statistical models also allow inferences about whether certain network substructures– often represented in the model by one or a small number of parameters – are more commonly observed in the network than might be expected by chance. We can then develop hypotheses about the social processes that might produce these structural properties.
- Sometimes, different social processes may make similar qualitative predictions about network structures and it is only through careful quantitative modeling that the differences in predictions can be evaluated.
- The more complex the network data structure, the more useful properly formulated models can be in achieving efficient representation. There are a variety of deterministic approaches for analyzing single binary networks, but many of these are not appropriate, or are too complex, for more complicated data.
- Several longstanding questions in social network analysis relate to the puzzle of how localized social processes and structures combine to form global

network patterns, and of whether such localized processes are sufficient to explain global network properties. It is difficult to investigate such questions without a model, as in all except rather simple cases the global outcomes resulting from the combinations of many small-scale structures are not immediately obvious, even qualitatively.

1.6 What SNA Does

There are a number of concepts and methods that are important to the study of social network analysis. Some of the most important ones include: social group, popularity, prestige, balance, transitivity, clique subgroup, social cohesion, social position, social role, reciprocity, mutuality, influence, conformity and dominance. It provides explicit formal statements and measures of social structural properties that might otherwise be explained only in figurative terms. Many of the expressions such as clique, popularity, isolation, prestige and so on are given mathematical definitions by social network analysis. Social network analysis has drawn on various branches of mathematics to clarify its concepts as well as to describe the consequences of its terms. It is deep rooted in mathematical graph theory. Moreover, social network analysis allows the measurement of structures and systems which would be almost impossible to describe without relational concepts, and provide investigations of hypotheses about these structural properties [3], [19].

Social networking analysis is not a formal theory but rather a broad strategy for investigating social structures [20]. It is grounded on the significance of relationships among actors in the network. Its primary objective is understand properties of the social structural environment and to show how these structural properties influence observed characteristics and associations among actors. SNA data require measurements on ties among social units. One cannot use multiple regression, t-tests, and structural equations to study social networks. SNA utilizes analytic techniques and applied statistics methods which usually focus on observational units and their characteristics.

The bottom line is that social network analysis can be studied and analyzed from a qualitative or quantitative perspective [21]. This is because networks have both a structure and process at the same time. Further, SNA provides insights on looking at a problem but does not predict what we will see. Finally, social network analysis embodies a range of theories relating types of observable social spaces and their relations to individual and group behavior using mathematical graph theories.

1.7 Why Use SNA

Social network analysis is motivated by the notion of, and built on the premise that information travels through contacts between actors, which can reflect a power distribution or influence attitudes and behaviors. Our understanding of social life improves if we account for this social space. Further, the patterns of contacts between actors can also effect the spread of information or power dynamics that could not be observed if we focus on the simple individual behavior.

Though the concept of social networks is very simple in nature, the amount of information we need to describe even small networks can be very complex [22]. To help managing the data and manipulate it so see the patterns of social structure can be highly tedious and complicated. These tasks can be readily done using mathematical tools or matrices. In many cases, the computer can be used to systematically analyze the data. In order to visualize the patterns in the network, graphs can be utilized.

Finally, social network analysts use mathematical graphs and matrices to represent patterns of ties among social actors. These concepts can be a daunting task to individuals without a mathematical background. Moreover, graphing that way can obscure various features of the social structure [22]. This may not be an efficient or practical way of visualizing the network. There are a number of useful network visualizing tools available for graphing your results. We used a few of these tools in this research to create a better picture of depicting the graphs. These include: UCINET; NetDraw; Pajek; igraph; Python; Rapid Miner; and KeyPlayer1.

1.8 Why Case Study?

We do not pretend to suggest that both of those cases present relatively new concepts in each of the different studies. Rather, the second case is an extension of the social network concepts presented in the first case study. Both cases highlight co-authorship and collaboration in social networks. It is important to note that co-authorship collaborative networks are not entirely unique. For example, the world wide web is an intricate evolving network with a constant change in the addition and deletion of nodes and links, resulting in overwhelmingly dynamic features [14], [17].

So why the case studies? This is because both co-authorship collaborative cases are explicit examples of dynamic evolving networks that share a similar structural characteristics. So in addition to inextricably having a map of the network topology, we also have a knowledge on when various nodes and links were added (or deleted) which is crucial for revealing the network dynamics.

1.9 Summary of dissertation cases

Each of the cases in this dissertation provides a different but varying perspective of some of the diverse methods used in detecting and interpreting patterns of social ties. The primary focus of this dissertation is to extend the study on SNA methods not only through the use of common everyday examples but also in areas that have seen little or no research.

Chapter Two provides the literature review for all the cases used in the study. Research has been done in social network analysis for decades. Our goal is to uncover what the researchers have found (or not found) in that area. We cover social networks and many of the concepts associated with social network analysis. This includes: degree distribution; social groups, popularity, prestige, balance, transitivity, clique subgroup, social cohesion, social position, social role, reciprocity, mutuality, influence, conformity and dominance, centrality measures, and so on.

Chapter Three examines the interactions and relationship between actors, groups and subgroups to ascertain the patterns, structure and attributes of the social network. The main concepts analyzed and discussed include the small-world theory, reciprocity, homophily, scale-free networks as well as power-law and power-law coefficient. This chapter extends on previous researches which have focused primarily on individuals running for political office but not on sitting government leaders.

Chapter Four is broken down into two phases. Phase One provides a more theoretical view of complex social graph analysis. The strength and the direction of the relationship – ties - in the network are examined. Using UCINET, Pajek, NetDraw, KeyPlayer and R, ties are evaluated to determine the existence of cliques, the degree centrality of the network actors, the cohesiveness of groups and subgroups, and the structural prestige of an actor from his or her social ties.

In Chapter Four Phase Two, another theoretical view is presented to analyze the intricate structure of the social network. Again, graph theory is used to examine the existence, the structure, and the presence of subgroups within the network. The primary SNA tools utilized to understand and evaluate the structure of the social networks include UCINET, Pajek, NetDraw, KeyPlayer, Rapid Miner, R and Python.

Chapter Five presents a summary of the essays and provides recommendations for future research in the subject area. We also examined the overall implications of the research and highlight possible limitations inherent with the studies.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

Online social networks such as Twitter and Facebook have increasingly become the desired platform for individuals to share information. The amount of information available can yield a treasure trove of data for social network analysis research.

Traditional SNA have focused largely on understanding the makeup and the structure of the social network. The primary objective of the traditional SNA is to calculate the relationship between nodes and links in a network and then analyze the degree of information flow through these nodes and links. Traditional SNA, often referred to as *static* SNA, has formed the basis of social network analysis. In recent years, however, much research interests have been concentrated on *dynamic* SNA.

Understanding social network structure and evolution have very significant implications for many aspects of network and design including defenses against computer attacks such as the Sybil attack [23], leveraging social networks for search, social regularization [24], information processing and diffusing social influence, bootstrapping trust via social networks [25], and provisioning [26]. In the past, many of the research work have focused on online social networks, scientific or biological network, and even predicting the outcome of a general elections. Our research have yet to come up with a study on elected officials' (in this case, governors) social network. The purpose of the first study is to use social network analysis to help us bridge that gap. In study one, our concentration is on static social network analysis. Specifically, we focus on the use of Twitter by state governors and study their interactions with other agencies, citizens, and other stake holders. Case study One and

Two are examples of dynamic social network analysis. Here we focus on the temporal dimensions of the social network analysis.

Since our study is motivated by social network analysis, the next section(s) contains a literature review of the theories, concepts and research on social network analysis.

2.12 Static Social Network Analysis

With static SNA, we study the state of a social graph $\boldsymbol{S}$ at a time $\boldsymbol{t}$ [27], [28]. It is built on graph models such as random graphs [9] and scale-free graphs [14] and measures which enables one to determine the implicit relationships among entities or information flow between entities in a network [29] by calculating degrees, connectivities, distances and flows. Essentially, the degree of a vertex $\boldsymbol{v}$ is a count of the number of edges connected to that vertex. By calculating the number of vertices that are accessible to $\boldsymbol{v}$, we are thereby determining its connectivity. In finding the minimal count of edges between vertices we are hereby finding the distance between these vertices. Therefore, the flow between vertices is a measure of the number of units flowing between these vertices [27]. Centrality measures are also key concepts of static SNA.

The degree of flow between vertices provide a basis for the centrality measures: *prestige; betweenness; closeness*; and so on. When a vertex is connected to a large number of vertices, either directly or indirectly, we say that that vertex hold some form of prestige. A vertex connected with a large count of close or neighbor vertices owns a high centrality of proximity. We can deduce that prestige and proximity are significant trust coefficient [27].

2.13 Dynamic Social Network Analysis

Social network analysis have focused traditionally on static approaches. However, it is important to note that many real-world social networks are not static in nature: they evolve with time. In the past few years, researchers have been interested in examining the evolution of these network communities to determine what triggers that expansion or contraction of these dynamic communities [30]–[33]. A dynamic community is defined as a collection of individuals who interact more frequently, contiguously and persistently among themselves than with other individuals [34], [35].

Dynamic SNA is used explicitly for modeling temporal changes in interactions. In dynamic networks, interactions are typically represented by a time-series of static networks, each network corresponding to interactions aggregated over a time frame [34]. Hence dynamic SNA examines the dynamic formation aspect of the social network, for instance how vertices are connected to other vertices with time, while at the same time, investigating the stochastic evolution of the individuals connecting the networks.

2.2.1 Government and web 2.0

Web 2.0 is a very broad concept. It can be viewed as a second generation of Internet content where the focus shifts from consumption to participation [36]. Web 2.0 can also be considered to be a network platform that delivers software and a service that dramatically improves the user experiences to that of Web 1.0 [37]. All in all, Web 2.0 technologies is closely associated with online collaboration, interactive information sharing, a design built with the user in mind and connection anywhere at any time [37]. Government agencies have started using Web 2.0 services to provide access to government services and/or increase interaction with citizens. *Table 1* and *Table 2* show a summary of some of the Web 2.0 tools used by the major law makers as well as previous literature on the use of Web 2.0 services by government agencies and officers in the United States. Here we focus just on services used for official business and those associated with permanent positions (like Speaker of the House, rather than the individual account of the current speaker).

As can be seen in from *Table 2,* there are several federal government agencies that use Web 2.0 technologies for various purposes. However, the primary focus appears to be information dissemination rather than to build social networks and to encourage public participation.

Table 1: Web Services Used By Major Law Makers

Government Title	*Web 2.0 Tools*
Speaker of House	Twitter, YouTube, Flickr, FaceBook, Speaker RSS Feed, Widgets to allow users to share information with friend on Twitter, FaceBook, Digg, delicious, StumbleUpon, Blogger, Tumblr, LinkedIn, Reddit
Department of Defense	RSS Feed, Twitter, FaceBook, YouTube, Flickr, DoDLive Blog, UStream, American Forces Widgets
Secretary of State	DipNoteBlog, Twitter, Tumblr, FaceBook, RSS Feed, Flickr, YouTube
Senate Majority Leader	Twitter, FaceBook, LinkedIn, Digg, Quora, YouTube
President	Twitter, YouTube, Flickr, FaceBook, MySpace, Vimeo, iTunes, LinkedIn, White House Blog, RSS Feed

Table 2: Use of web 2.0 technologies by government agencies and officials

Agency	*Web 2.0 Service*	*Tools*	*Aim*
General Services Administration (GSA) [38]	Mashups, Web services	Usa.gov, GobiernoUSA.gov, GovGab.gov, kids.gov, Webcontent.gov	Develop strategy for government agencies to provide better services using Web 2.0
Federal Bureau of Investigation [38]	Mashups (widgets)	FBI Most Wanted Widgets	Provide capability for other government sites to add the FBI most wanted list
DoD & other Intelligence Agencies [38] [39]	Wikis, Mashups, RSS feeds, Rich Internet Applications	Intellipedia, DoDLive, Virtual Worlds, TroopTube	To improve intelligence sharing, contribute and sharing of content using simple markups
The White House [40]	Podcasts	White House podcasts	Provide updates, coverage of live government deliberations, emergency response information
state Department, Environmental Protection Agency (EPA), Transportation and Security Administration (TSA) [40]	Social Networking – Facebook, Twitter	GreenVersations	Support public interaction in response to agency announcements
Center for Disease Control [41]	RSS feeds, Instant Messaging, Podcast	Flu Wiki, Second Life	Provide emergency text messages, seasonal flu updates
Library of Congress [41]	Multimedia Sharing		Distribution of digital content
National Aeronautics and Space Administration [40] [39]	RSS feeds, Social Networking	SpaceBook, Twitter	Publicize events, news releases, real-time updates and information from space stations
Department of Education [38]	Mashups (Facebook, Twitter widgets)	College Navigator, Federal Student Financial Aid ForeCaster	Make college more accessible and affordable
state Department [40] [39]	Wikis, blog, Social Networking	Diplopedia, Deskipedia, communities@state, Exchanges Connect, DipNote	Internal foreign affairs encyclopedia; blogging platform organized into communities of practice and interest; tool similar to Facebook
Army [40] [39]	Wikis, blog	MilSuite, MilBlog, MilWiki, MilBook	A series of tools to disseminate information internally
Joint Staff (JS) [40] [39]	Mashups (widgets), Instant Messaging	All Partners Access Network (APAN), Intelink, Jabber	Network to foster interagency collaboration and coordination; chat rooms in multiple networks that help speed up information gathering

United states Geological Survey (USGS) [40] [39]	Mashups	Twitter Earthquake Detector Project	Provides the USGS with initial indication of an earthquake before the scientific data reaches the USGS
Department of Education	Rich Internet Application	ED Data Express	Interactive web site aimed at making timely and accurate K-12 data available to the public
Housing and Urban Development (HUD) [40] [39]	Mashups	USA Search, HUD National Housing Locator System	Provides a searchable web based database of available rental housing nation-wide
Treasury/IRS [38]	Web Service	IRS eFile	Make it easier for taxpayers to pay taxes quickly and accurately

2.2.2 Social networking in government

In this study, we focus on the use of Twitter by state governors and study their interactions with other agencies, citizens, and stake holders. Posts on Twitter (known as 'Tweets') allow Twitter users to update and share information readily with individuals who follow them. People who receive or subscribe to your tweets are referred to as your followers. Your friends (often called followings) are other Twitter users you have chosen to follow. A one-way or two-way relationship may exist between followers and followings, however, unlike most other online social networking sites, Twitter does not require any level of reciprocity between followers and friends [42]. Another interesting functionality of Twitter is 're-tweeting.' Re-tweeting is the Twitter-equivalence of email-forwarding, where users post messages which were originally posted by other Twitter users [43]. Using the power of re-tweeting, social micro-blogging sites like Twitter have enabled individuals, groups and organizations to broadcast, share and disseminate information about their activities, opinions, and status remarkably easily through their network of followers or friends [42], [44]. Among young adults 18 – 24 years old, about thirty-seven percent make or read Twitter online updates [45]. Twitter is free, it is relatively low network resource intensive, it is easy to learn and use, and integration with mobile services and other existing applications is very simple. These features give it the potential to radically extend the communications reach and make it a viable option for the adoption by the government [38], [46], [47]

Table 3 summarizes a selected set of studies on the use of Twitter and other similar social networking services in the government. Though these studies have interesting findings, none of the studies have looked at the structure of the networks of government agencies and citizens. To understand the use of Twitter by state governors, we look at two primary characteristics:

- What is the level of interaction in the network – here we look at the reciprocity in the network to see if the governors are also following the citizens and other agencies in their networks

- What are the spatial characteristics of the network of followers – here we look at geographical homophily and examine the geographical spread of the governors' networks

Table 3: Previous studies on the use of social networking tools by the government

	Aim	*Finding*
[48]	Understanding how to leverage web 2.0 for Government-citizen and government- employee interactions	Among the findings include: Government needs to meet citizens where they are online; citizens are willing to interact with government agencies online; etc.
[38]	Author examined the role of Twitter in government agencies	Government organizations are using Twitter to disseminate information both internally and externally
[46]	Research to explore the adoption of Twitter by government agencies	Twitter is used extensively for conversation and collaboration in projects
[49]	Research questions revolve around the significance of web 2.0 for e-Government	Web 2.0 presents significant opportunities as well as risks for government
[50]	Study to determine to what extent citizens are using government online services and information	Americans are turning in large numbers to access online government information and services
[51]	Study to examine the Twitter accounts of federal and state politicians to determine how they are using this social networking tool	Government agencies primarily relied on a one-way communication that sought to inform and communicate rather than two-way symmetrical communications
[41]	Among other things, to determine how government agencies are using web 2.0 tools	Web 2.0 tools are already being deployed internally through government agencies.
[52]	Determine the type of content legislators are posting	Law makers are primarily using Twitter to post information to their constituents

2.2.3 Reciprocity

Social networking typically involves a high level of reciprocity which can be defined as a pattern of mutual exchange between actors [3]. Ever since the inception of the social media and networking sites and their popularity, there has been a renewed interest in the reciprocity of social networking [53]–[55]. The typical new users of social networking are added by mutual acquaintances or because their friends invited them to do so. This generally evolves into bi-directional/reciprocal communication between friends. However, Twitter networks are slightly different as they have two kinds of relationships - followers and friends. Friends generally follow each other to share information or to keep abreast of new developments in each other's lives. However, that may not be the case for followers. We investigate this phenomenon in the networks of state governors to examine if they follow the accounts of their followers.

Some studies have been done to investigate reciprocity in Twitter. Java et al. (2007) compared micro-blogging to regular blogging and found that users of micro-blogging social networks have a significantly high degree of correlation and reciprocity [56]. They determined a 58% reciprocity in Twitter friends – which mean that 58% of the users were likely to follow the users who were following them. Kwak et al. (2010) found that only about 22 percent of Twitter users have some form of reciprocal relationships. They also observed that 67% of users are not followed by their friends. Other studies on social networks such as Flickr [16] and Yahoo [57] have found the reciprocity to be 68% and 84% respectively.

2.2.4 Geographical homophily

Sociologists have found that individuals who live in a common geographic area, have work or family ties or have common interests have the tendency to relate to each other [58]. This phenomenon is called Homophily and is defined as the principle that contact between similar people occurs at a higher rate than among dissimilar people [59]. Essentially, homophily implies that the "distance in terms of social characteristics translate into network distance" [59]. Geographic homophily is the most fundamental of all homophilies. Contacts occur more frequently with individuals who are in close proximity with the other. This is because it takes a lot more effort to establish and maintain relationships with someone who is far away than one who is readily available. With the proliferation of online social networks, one would deduce that spatial and temporal factors have a lesser effect on the establishment of connections between people. A recent study by Kwak et al. (2010) found that Twitter users who have reciprocal relations tend to be geographically close. In general, it was found that a slight homophily exists only if there is some form of reciprocity [42]. A study of a sample of users of LiveJournal [58] found that geography also affects homophily. They concluded that two arbitrary users, even with a remotely few common interests, are more likely to be friends especially when not restricted by organization or geographic constraints such as distance and time [59].

2.3.1 Social structure

Social structures can be referred to as enduring patterns of behavior and relationship within social systems or to social institutions and norms that have been embedded into that social system and which is some way shape the actions of that individual within that structure [60]. The perception that social structure can be deemed as relationships between entities and groups or as enduring patterns of relationships underscores the concept that society is grouped into structurally related groups or sets of roles with different meaning, purposes, and functions[60]. A social network is a social structure made up of individuals, organizations or nations, commonly called actors or nodes, and the relationships between those actors, dubbed “ties”. Social network analysis (SNA) offers the methodology to analyze social relations present in a social network. In other words, SNA tells us how to conceptualize social networks and further, how to analyze those social networks[5].

Social network analysis assumes that interpersonal ties among actors transmit behavior, attitudes, information or goods[5]. In fact, the primary objective of SNA is detecting and interpreting patterns of social ties among actors. Moreover, SNA also assumes that the structural location of an actor has significant perceptual, behavioral and attitudinal implications. This is because the location of the actor within the social network impacts our beliefs, norms and observed behavior in the social network[61].

The graph theory is used to analyze a social network. A graph is considered to be a set of interconnected objects represented by vertices and a set of lines between pairs of vertices, called edges. Lines are directed or undirected. Whereas a directed line is called an arc, an undirected line is called an edge. An arc points from a sender to a receiver. On the other hand, an edge has no direction and is represented by an unordered pair. Unlike directed graphs that contain one or more arcs, undirected graphs contain no arcs: all lines are edges[5].

Despite the fact that many researchers have categorized social network analysis as more of a method than a theory[62], the majority of social network analysis methods are developed from the following set of assumptions [3], [63]:

- Actors and their related actions are interdependent, rather than independent, with other actors
- Ties between actors are seen as channels for the transfer or flow of various types of resources (e.g., funds, information, trust, enmity, etc.)
- Social structures are seen in terms of enduring patterns of ties between actors
- An actor's position in the social structure (i.e., its structural location) impacts its beliefs, norms and observed behavior

- Social networks are dynamic entities that change as actors, subgroups, and ties between actors enter or leave the network.

2.3.2 Degree distribution

Degree distribution is by far one of the most important statistical characteristic of any network[64]. Although social network analysis is not graph theory, there are fundamental concepts in graph theory that can be used and are being used to describe social network analysis. In graph theory as well as network analysis, the degree or connectivity of a node is the relative number of edges the node has to other nodes. The degree distribution ***P(k)*** of a network can be defined as the probability that selected nodes in the network has ***k*** edges[65].

Social networks exhibit a full spectrum of degree distributions ranging from one extreme where the distribution of links is nearly as if they were formed uniformly at random and another extreme where most nodes have only a few links, held together by a few highly connected hubs[64]. With a random network, a researcher will have roughly the same number of collaborators and the resulting distribution will be a bell-shaped Poisson curve that peaks at the average number of collaborators[66]. It is less likely to find groups in random networks. Needless to say, random network models are not able to capture important features of many observed social networks, for example, the combination of relatively small diameters and high levels of clustering[8] . On the other hand, small-world networks[7], despite often their large size, have a relatively short path between any two nodes.

2.3.3 Cohesive sub-groups

Within the larger network structures are embedded groups of actors who interact with each other. These dense clusters of actors are referred to as cohesive subgroups. One of the primary focus of social network analysis is to identify dense clusters of actors who share strong, direct, frequent or positive ties[3]. Generally, groups or subgroups are formed because of a feeling of belongingness for each other, individuals have similar beliefs or values, or perhaps individuals exhibit a collective behavior that link them together.

2.3.4 Centrality measures

Social networks are likely to comprise of people or organizations that are central in the network structure. Wasserman and Faust (1994) argued that the most important uses of graph theory in SNA is to identify the most central actor in the network. The concept that certain actors in a network are more central than others can be traced as far back as Moreno’s conception of sociometric stars and isolates[5]. In the 1950s, Alex Bavelas and his team of MIT researchers were the first to formally investigate the properties of centrality as he looked at how information travelled effectively and efficiently within network structures [3], [4], [61]. Bavelas argued that inherent with their location; central actors can influence the flow of communication within the network. A multiplicity of measures, each based on different assumptions of what it means to be central can be used in social network analysis. Some of the most commonly used ones include: degree and betweenness.

- Degree centrality is the measure of the number of immediate ties an actor has in a network. With degree centrality, the direction of the ties is not measured.

Rather, it is used to measure an actor's level of involvement, activity or immediate influence in a network [4], [67]. Degree centrality for a node or actor can be calculated:

$$C_D(p_i) = \sum_{k=1}^{N} a(p_i, p_k)$$

where, $a(p_i, p_k)$ = 1, if there is a direct tie between p_i and p_k and i ≠ k [68].

- Betweenness centrality measures the extent to which each actor lies on the shortest path between all other actors in a network. It differs from degree centrality in that it assumes that an actor has power over any two other actors when it lies on the shortest path between them in a given network of relations. Betweenness centrality measures how much potential control an actor has over information flow within the network. For example, if an actor lies between many other actors in the network, then the actor can greatly influence the network by choosing to withhold or distort information he or she receives[4]. Betweenness can be calculated using this formula:

of a node can be given as: $$\sum_i \sum_j \frac{g_{ikj}}{g_{ij}}, \quad i \neq j \neq k$$

where g_{ij} is the number of geodesic paths from i to j, and g_{ikj} is the number of these geodesics that pass through node k [67].

2.3.5 Popularity

In social networks, people who receive many positive choices have some form of prestige in the network. The popularity, also referred to as indegree of a vertex is the number of arcs it receives in a directed network. Popularity is a measure of structural prestige: prestige based on a person's social ties. A person's nominations on a social positive relation suggest a sign of prestige. However, if he or she continues to receive more nominations it indicates higher structural prestige[5]. In other words, an actor with a small indegree is chosen by few others, while an actor with a large indegree is one in whom many has nominated as friends[3]. An actor who holds a position of great popularity is an ideal position to assume leadership roles and influence information flow.

2.3.6 Cliques, strong and weak ties

A clique is the strictest structural form of a cohesive subgroup. It is a set of vertices in which each actor is directly connected to all other actors. In other words, a clique is a subnetwork with maximum density. Technically, cliques ideally contain a minimum of three actors because although smaller subnetworks exist, they are relatively uninteresting because subnetworks of size 1 and 2 are single vertices and edges or bidirectional arcs respectively. Detecting cliques in large networks can be a time-consuming process because even medium-sized networks often contain a large number of cliques. Certainly, sometimes you will find more cliques than actors. In social network analysis, structures of overlapping cliques (that is triads share one or more vertices), often considered as social circles than individual cliques are regarded as cohesive subgroups. As cliques evolve the subgroups develop its own set of norms, rules and culture that are different from the network it is embedded. Consequently, such cliques can be used as a reference points for individuals and individual's identity. On the other hand, overlapping cliques can cause confusion in interpreting the results of a clique analysis because it hides the underlying clique structure [3]–[5], [61].

Within a social network, actors are linked together by ties which fluctuate in both their type and strength. Ties differ on a continuum from strong to weak[62]. At the individual level, we can think of strong ties as those where actors have repeated and relatively intense interactions with one another, whereas we can think of weak ties are actors who see one another occasionally or rarely[61]. Ties have many implications in the study of social networks operations, information flow as well as how structural nodes can play structurally distinct roles in the information diffusion process. For example, an analysis of the March 11, 2004, Madrid bombings, Rodriguez[69] found that weak ties were a key feature of the terrorist network in that they enabled its cells to maintain operative ties with the larger network from which they were able to draw material supplies and ideological support. Moreover, Rodriguez believes that weak ties provide benefits to dark networks in other ways. He argues, that weak ties provide dark networks with: relative stability when members are arrested or missions fail; more flexibility that allows them to rapidly adapt to a changing environment and (3) higher levels of security because weak ties are harder to detect than strong ones[61].

The above statement does not suggest that strong ties are of no value. There is undeniably, a vast amount of research supporting the fact that people with strong ties are happier and even healthier because in such networks members provide one another with strong emotional and material support in times of grief or trouble and someone with whom to share life’s joys and triumphs. Therefore, feelings of trust and solidarity are more likely to be shared across strong ties than across weak ones[61].

2.3.6 Proximity and structural prestige

The choice of a maximum distance from neighbors within a restricted input domain is quite arbitrary. Proximity prestige overcomes this problem by considering all vertices within the input domain of a vertex but attaching more importance to a nomination if it is expressed by a closer neighbor. That is, nomination by a close neighbor contributes more to the proximity prestige of an actor than a nomination by a distant neighbor. To allow direct choices to contribute more to the prestige of a vertex that indirect choice, proximity prestige weights each choice by its path distance to the vertex. Hence a higher distance yields a lower contribution to the proximity prestige. Proximity prestige of a vertex is the proportion of all vertices in its input domain divided by the mean distance from all vertices in its input domain[5], [61]. Finally, proximity prestige can be directly correlated to rank and social status of a social network.

2.3.7 Diffusion

Information diffusion is one of the fundamental aspects of social network analysis. Social ties allow one to access information which can be used to reduce uncertainty and risk and to create trust within the network and subnetwork. This is absolutely important because people in crucial positions in the information network may also strategically spread and retain information because they have control over the diffusion of that information. In a social system, sometimes the overall structure of the informal ties can both directly and indirectly impact the level of information diffusion throughout the network. For instance, in a network structure where individual actors or subgroups pursue their own agendas, this could inhibit information flow, thereby resulting in a network bottleneck.

2.4.1 Patterns of Interaction in Social Networks

Networks, in particular social networks, have continued to attract the interests of researchers for years [70]–[72]. In recent times, however, the keen interest in social networks and rather, social network analysis, is sparked primarily to its relevance to social, and dynamic processes, as well as the diffusion of social influence in the network itself [71]. The formation of a social network is a highly complex process which comprises of a substantial number of characteristics. For example, the structure of a network consists of lines that link a vertex. Several of these vertices may be connected to form arcs [4], [5]. A *vertex* is analogous to an actor and is considered to be the smallest unit in a network. The goal of *social network analysis* is to detect and interpret the patterns of social ties in a network [5].

When the actors in a network interact, links are formed. The strength of these links measures the importance of the node [62], [73], [74]. These affect the information diffusion throughout the network. Hence, *strong ties* tend to bond similar people together to form clusters whereas *weak ties* are like "local bridges" that connect parts of social system that are otherwise disconnected [62], [75]. Sometimes actors in the social network may develop shared norms, identity, solidarity, and even cliques, as a result of this social cohesion [4], [5].

Many of the possible interactions in a social network are generally random in nature. Researchers have come up with various structural models to represent these social networks including: *the small-world model*; *preferential attachment model* and the *scale-free model*. A small-world network can be described as a network in which the level of local clustering is significantly high while the proportion of geodesics between pairs of vertices is relatively small [7], [8]. Barabasi, Albert and Jeong found that large networks are self-organized in a scale-free state in which a *power-law* distribution is likely arise and no single state characteristics can be defined [14], [76].Theoretically, this view holds that star connectors or collaborators are crucial to the development and integration of these kinds of network. The preferential attachment model describes a growth network model in which new nodes attach preferentially to old nodes that are already well established or well-connected [14], [77]. Essentially, preferential attachment is a variant of the popular saying, “the rich get richer” [5].

2.4.2 Social Network Matrices – Closeness, Betweenness, Structural Cohesion

Centrality refers to the position of the vertex in a network. The degree of centrality is the distance of an actor from other actors in the network. Those actors that are most centrally located are optimally positioned for integrating information throughout the network [6]. Distance is a key strategic variable in social network analysis because it affects information diffusion throughout the network. Other indices of centrality include *closeness* and *betweenness*. The *closeness* centrality is a measure of how close an actor is to all other actors in the network. Technically, it measures the shortest path to all actors where the larger distances yield lower closeness. Closeness affects information dissemination because the closer a vertex is to all other vertices(high centrality), the easier the information may reach it [5], [73], [78], [79]. Closeness centrality for a node or actor can be calculated:

$$C_c(p_i) = \frac{N-1}{\sum_{k-1}^{N} d(p_i, p_k)}$$

where, N is the number of node or actors in a network and i ≠ k [68].

Betweenness is a measure of the number of the geodesic paths that pass through a node. In other words, it is the most likely channel for transporting information through the network. The more pairs of vertices that pass through the actor the more central that actor is [19], [67], [79], [80]. The betweenness centrality of a node can be given as:

$$\sum_i \sum_j \frac{g_{ikj}}{g_{ij}}, \quad i \neq j \neq k$$

where g_{ij} is the number of geodesic paths from *i* to *j*, and g_{ikj} is the number of these geodesics that pass through node *k* [67].

Sometimes it becomes very important to determine what is the minimum number of actors, who if removed from the group would disconnect the group. We refer to this as *structural cohesion*. Structural cohesion is founded on homogeneity and explains the reasons behind subgroups within groups; and cliques within groups or subgroups [81], [82].

2.4.3 Network Evolution

With social network analysis, much of the focus has been on the evolution of the social network. The two primary mechanisms for understanding network evolution are *preferential attachment* and *homophily* [14], [77], [83] . Preferential attachment argues that new nodes will attach preferentially to already well-connected nodes.

Preferential attachment implies popularity. Homophily, on the other hand, is the principle that contact between similar people occur at a much higher rate than dissimilar individuals [83]–[86]. Homophily is a variant of the analogy "birds of a feather flock together". It implies that people who have a common status or perhaps a common value are more likely to collaborate than those where these factors are different.

Empirical evidence have shown that the degree distribution of the nodes of large social networks follows a *power-law* which significantly influences the robustness of the network [65], [87], [88]. That is, the links of some social networks quantify some level of significance or weight. Uncovering your weak links, for instance, can determine your "bottlenecks" [87].We use the *minimum spanning tree* (MST) to uncover the location and role of weak links in a complex network [87].

2.4.4 Useful Social Network Analysis Formulas

Below are some useful formulas that can be used to calculate the some of the concepts of the network evolution concepts:

a) Preferential Attachment

The probability $\pi(\mathbf{k})$ that one of the links of the new node connects to node i depends on the degree $\mathbf{k}_i$ of node i as [89]:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

b) Power-law

A quantity x obeys a power law if it is drawn from a probability distribution

$$p(x) \propto x^{-\alpha}$$

where α is a constant parameter of the distribution known as the exponent or scaling parameter; most scaling parameters lie in the range $2 < \alpha < 3$ [90].

c) Clustering Coefficient

The clustering coefficient **C(G)** of a graph **G** is the average over the clustering coefficients of its nodes

$$C(G) = \frac{1}{|V'|} \sum_{v \in V'} c(v),$$

where **V 0** is the set of nodes **v** with **d(v)** ≥ 2 [91], [92].

d) Cumulative Distribution Function

The cumulative distribution function ***F(x)*** for a continuous random variable ***X*** with probability density function ***f*** is defined by:

$$F(x) = P(X \leq x)$$

The cumulative distribution function ***F(x)*** for a continuous random variable ***X*** with probability density function ***f*** is defined by:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\, dy$$

e) Complementary Cumulative Distribution Function

$$\overline{F}(x) = P(X > x) = 1 - F(x) = \int_{x}^{\infty} f(x)\, dx$$

CHAPTER THREE: STATE GOVERNORS ON SOCIAL MEDIA: RECIPROCITY AND HOMOPHILY IN TWITTER NETWORKS

3.1 Introduction

Government transparency and the right to government information, among other things, are highly regarded as fundamental to democratic participation and informed decision making by the government [93], [94]. With the advent of the highly participatory form of the web (commonly known as Web 2.0), social networking and other services can be used to communicate both within government agencies and externally to citizens [46] as well as to build relationships with stakeholders.

Since the 2008 general elections, social media has increasingly become a cradle for political activities and activism in the United States. Web 2.0 appears to be a good platform for both political communication between citizens and for communication between candidates and their constituents [49]. It has been suggested that the unprecedented speed in which changes are occurring in the Middle East was due to the social media and social networks [95], [96]. Research studies have also found that predicting the outcome of general elections by analyzing the popularity meter and sentiments for the candidates may be possible [97], [98]. While there has been much study on how political candidates use social networking tools [36], [49], [97], [99], [100], there has been little work on the use of these tools by government officials, government agencies, and other stakeholders interacting with the current administration. Though there has been a significant increase in the number of online services provided by government agencies, social networking sites like Facebook and

Twitter are severely underutilized and it has been suggested that agencies should use those tools more readily to reach the under-served populations [38], [46], [48], [50], [52].

In this study, we focus on the use of Twitter by state governors and study their interactions with other agencies, citizens, and other stake holders. We explore the following questions: how are government agencies using the social networking media to communicate with its citizens? What are the properties of their social network? Who are the agencies following and is there a degree of reciprocity between followers and friends?

3.3 Research design

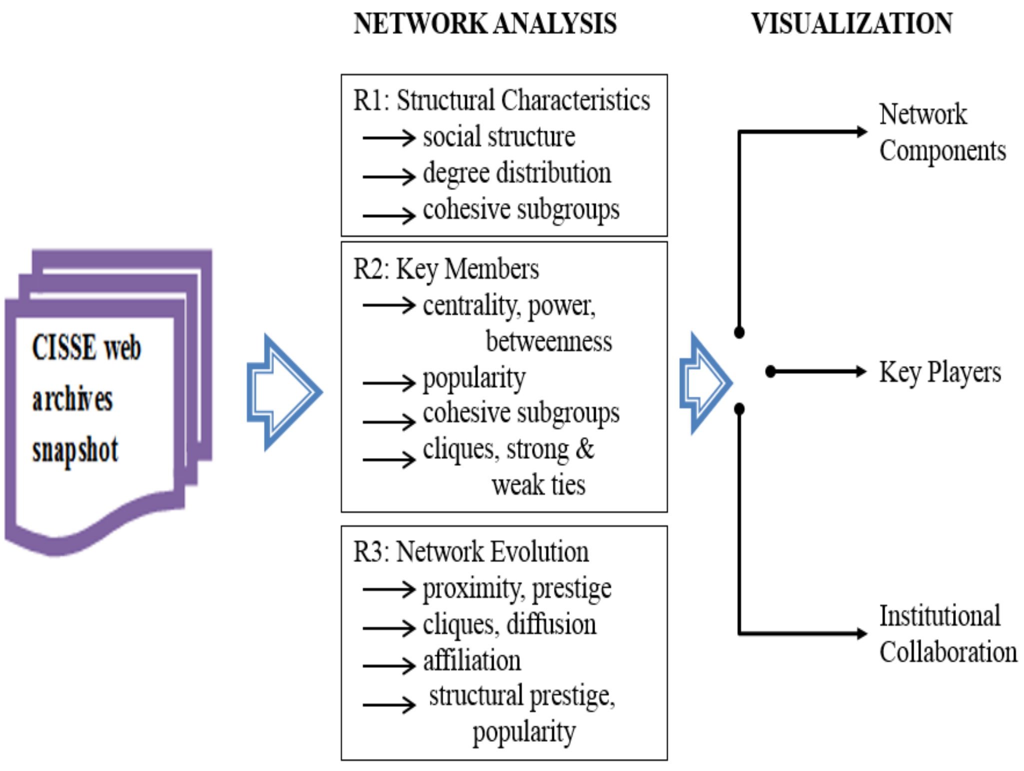

Figure 1: Research design process

3.3.1 Data acquisition

To select state governors who are active on Twitter, we gathered data on 10 US state governors with the most number of Twitter followers and friends. Twitter keeps a public profile of each user which includes the name, a brief description, screen name (pseudonym), location, the total followers and friends count. Using the Twitter API (https://dev.twitter.com/), we downloaded a two-level network for each governor using the friend and follower relationships. For example, the governor was at level zero, the friends of the governor were at level 1; their friends were at level 2. To work with the download limit of Twitter, we used a self-regulating, automatic download process which allowed us to gather 3,500 users per hour. Between March, 2011 and November, 2011 we collected information for a total of 3,892,868 users (5,470,647 as followers and 4,133,087 as friends).

3.3.2 Analysis

In this section we report the analysis outcomes and answer our research questions about the level of interaction in the governors' network to determine the existence or not of reciprocal relationships within the network; the spatial characteristics of the network of followers for geographical homophily and the spread of graphical spread of the governors' network and; the structural characteristics of the State governors network.

3.3.2.1 Power Law

When data is presented pictorially on a graph, one can easily see the salient features in the data and could thus interpret it accordingly. If the data is evenly distributed around the mean value, we commonly refer to this characteristic as a normal curve, bell-shaped curve, or normal distribution. But this is not always the case. Sometimes, the distribution may have more sample data with extreme values than normal distributions, drawing a curve with a long tail lowering as the value increases. Such distribution characteristic is called a power-law distribution. The shape of the power-law distribution has precipitated a number of buzzwords including 80/20 Rule and the Winner Take-All Society [101]. Research has shown that many economic, physics, biology, geography and sociology phenomena, as well as social networks follow a power-law distribution.

Vilfredo Pareto observed in Italy that 20% of the population at that time was holding 80% of the wealth. This was later coined the Pareto principle which states that 80% of consequences stem from 20% of causes. George Zipf also observed that word use frequency falls in a power-law pattern, with a small number of high frequency words, a moderate number of common words and a large number of low frequency words. Shirky argued that in any system where people are free to choose among alternatives, a power-law distribution would inevitably be created [101]. Albert and Barabasi posit that the vertex connections of many large networks, including the World Wide Web, follow a scale-free power-law distribution [102]. They concluded that “a vertex that acquires more connections than another one will increase its connectivity at a higher rate; thus, an initial difference in the connectivity between two vertices will increase further as the networks grow [102]. Essentially, Barabasi and Albert postulate that older vertices will increase their connectivity at the expense of younger ones leading to highly connected vertices over time. Adamic et al. deemed Barabasi’s theory inconsistent with empirically observed properties of the World Wide Web network topology [15]. In their research, Adamic et al. found no correlation between the age of a site and its number of links. Conversely, they found that the more appealing a site is, the more visible it becomes. Work on Twitter and other micro-blogging services has been inclusive on the distribution of links among friends and followers, some studies [103], [104] found a power-law following distribution and Kwak et al. looking at a large network [42] concluded that a non-power-law distribution exists in Twitter follower-following topology.

3.3.2.2 Reciprocity

Reciprocity is an indicator that a symbiotic relationship exists between a user and her /his friends and followers. In Twitter, a reciprocal follower relationship exists between user *U* and user *V* if *U* follows *V* and *V* follows them back. Thus, if *U* posts a message on Twitter then *V* sees it and vice-versa. The same applies for friendship relations. For each governor, we looked at the reciprocity at the first level of the network to determine the number and percent of friends and followers that the governors reciprocated with. A high level of reciprocation by the governor would imply that he/she can see the tweets posted by the citizens or agencies in their network.

3.3.2.3 Geographical homophily

Twitter users contain a location attribute in their profiles which can be used to establish the geographic homophily of each of governor's followers. In addition, using this variable, we can establish to what degree the governors' followers are located within its state, nationally, internationally or within the proximity of its ten largest cities. Since e-Government is targeted towards providing services for its citizens, it is therefore anticipated that the majority of the followers for any governor should be located within its state, however with online social networking, distance may be a lesser factor in determining relationship.

3.4 Experimental results and discussion

Table 4 shows the basic statistics for the 10 U.S. state governors and their first and second level friends and followers. The governors have significantly disproportionate number of Twitter followers than friends. A main reason for this could be that governors overall use Twitter as a means of disseminating information to as well as communicating directly with its citizens rather than collecting information on what its people are doing.

Table 4: Basic statistics of the dataset

Governor, State	*Followers 1st Level*	*Followers 2nd Level*	*Friends 1st Level*	*Friends 2nd Level*
Jerry Brown, CA	1,090,111	4,446,441	4,093	4, 445, 595
Deval Patrick, MA	19,505	1,393,940	11,524	1, 384 ,267
Bill Haslam, TN	8,118	73,375	2,959	605, 806
Martin O'Malley, MD	6,788	97, 583	823	96, 757
Chris Gregoire, WA	5, 748	199,341	24	199, 326
Nathan Deal, GA	5,693	46,367	15	46, 350
Andrew Cuomo, NY	4,834	13,368	143	105, 278
John Kitzhaber, OR	3,093	170,558	1,688	169 ,062
Brian Sandoval, NV	2, 731	39, 975	1,959	38, 016
John Hickenlooper, CO	2,669	374,439	993	373, 594

3.4.1 Power Law

From the onset, we did a topological basic analysis of the characteristics of friends and follower. *Figure 2* below provides a breakdown of followers and friends count for governors at the first, second and third network levels, respectively. In *Figure 2A* to *Figure 2C*, we explore the probability that a random variable or value **N** with a given probability distribution is less than or equal to **N**. In order words, all three graphs were developed using the Complementary Cumulative Distribution Function (CCDF).

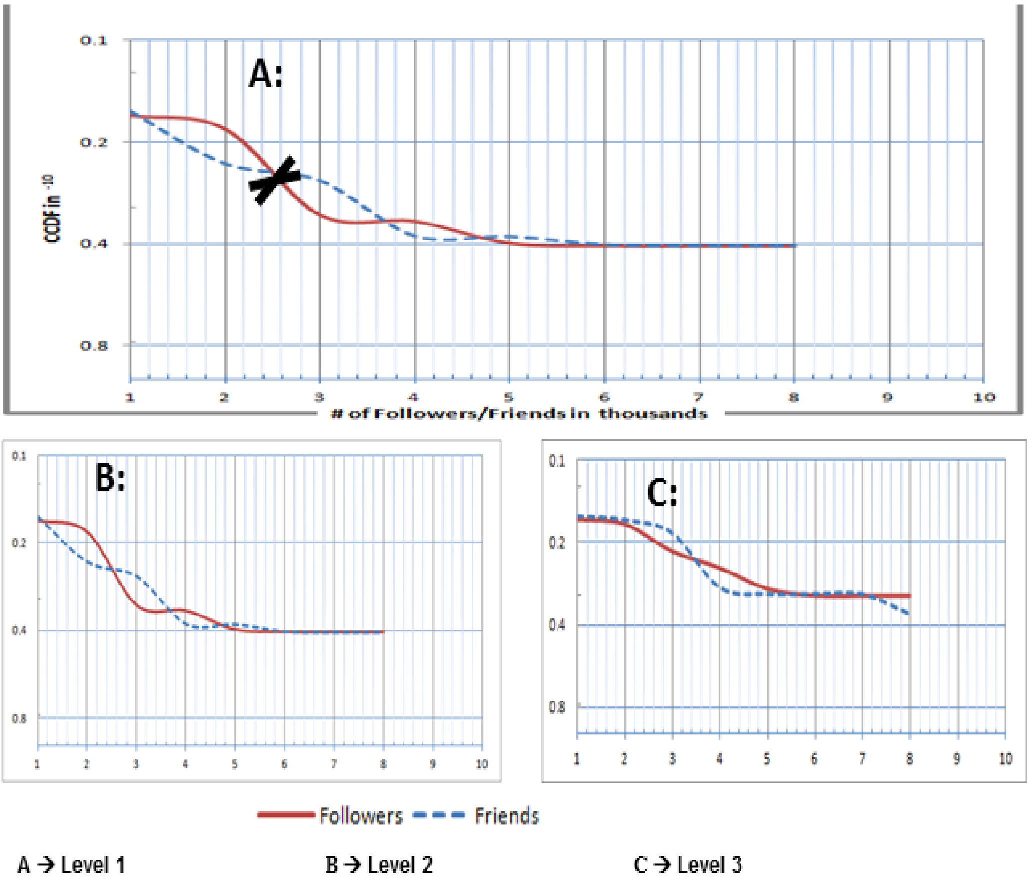

Figure 2: Followers and Friends of US State Governors

Figure 2 measures how many users are at or above the given follower or friend level. The followers or friend level is expressed in thousands of followers or friends. The **X**-axis represents the number of followers and friends in thousands of users. The Y-axis is the percent of time the followers and friend count is at or above the number of followers or friends specified by the **X**-axis. For example, the **X** in *Figure 2A* represents a follower or friend level where both friends and followers are relatively identical with a count of 2,500 and a percentage of negative thirty percent. It can be observed from the graph that the number of Governor followers from 1 to 2,500 at the first level is much higher than for the friend count at that same level.

Further, it can be observed from *Figure 2* the number of friends and followers at the first level of the network fluctuate at an alternating pattern until the number of friends and followers approaches 5000. The same is true for level two. However, for level three, the undulating pattern is less pronounced. Further, the graphs for *Figure 2*A - C show that at 103 or between 1,000 and 10,000, the number of users is larger for friends than for followers. Generally, as can be examined from the graphs, very few users in the study have friends in excess of 5000. On the other hand, a sizeable number of users have more than one million followers.

Figure 3 below illustrates that at level one, a power law exists for Governor followers at **x** = 1262 with a power exponent of 7.14. The power law and power exponent for Governor friends are 1075 and 6.98, respectively. Our power law result differs significantly from previous work done on social network sites, in which researchers found that the normal power law exponent for most real networks to be between 2 and 3 [3], [96]. The results of our study, therefore, suggest that the power law exponent for friends is marginally stronger for Twitter friends than it is for followers. Further, judging by the high exponent value in the research, one can potentially say that a power law may not exist at all.

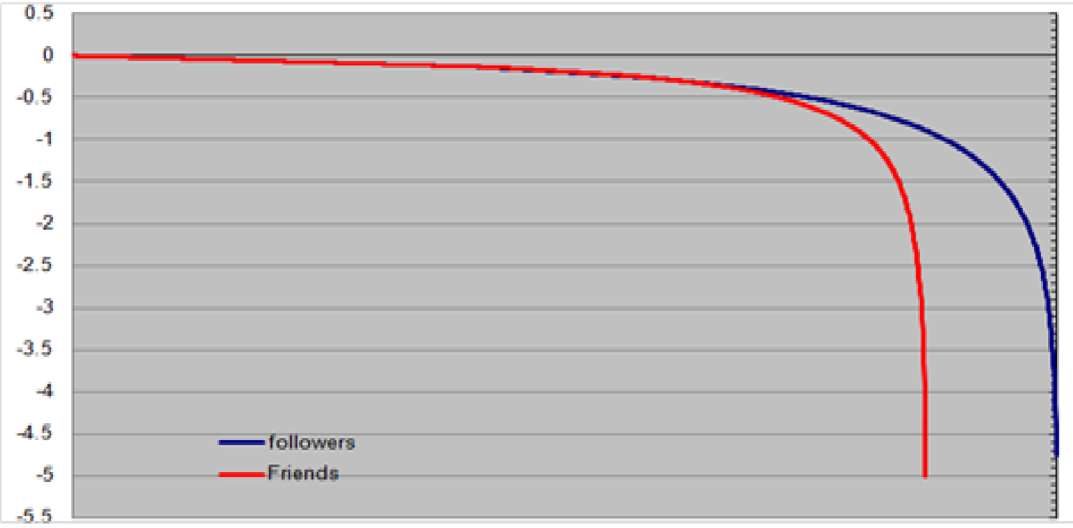

Figure 3: Level One Followers/Friends Power Law for O'Malley, MD

Table 5 below shows the correlation between followers and friends at either levels of the network. In our study, we observed that about nine percent of the followers in the network were also found in the friends' network. We also sought to determine whether the number of followers at each level of the network had a direct or inverse correlation with the number of friends at each level in the governors' network. *Table 6* shows the results of the Pearson Correlation Coefficient between the friends and followers at each of the levels and established the mean of the distribution to be 0.7683. It can also be observed from *Table 6* that at the third level an increasing number of followers count is almost directly proportional to an increase in the number of followers. Based on these results, it is difficult to surmise the rational for the very low correlation coefficient at the second level.

Table 5: Friends that are also Followers

Total # of Friends	Friends that are also Followers	Percentage Friend that are followers
1,418,724	115,372	8.13%

Table 6: Correlation between Followers and Friends

Level	Correlation
1	0.88348
2	0.50576
3	0.91585

3.4.2 Reciprocity

With over five million followers and four million friends in the study, there are only twenty-nine instances of a governor following a second-level user. Conversely, if a User A follows a governor B, and User A is followed by User C, then there were only 218 (0.003 %) cases at the second level that that User C followed governor B. On the other hand, that number was 724 (0.002%) for friends. Thus, the network for governors does not appear to show the characteristics of more typical friendship based social networks [16] and shows much lower levels of reciprocity than other more general Twitter networks [42].

Table 7 shows the details for all the governors. It can be observed that the followers' reciprocal count for each of the governors is considerably smaller than that for friends. The reciprocal percent is a measure of the ratio of the reciprocal count for that governor at that particular network level to the total number of followers or friends at that network level. It can also be seen that the reciprocal percentage of friends for each of the governors is several times higher than for followers. These low reciprocal percentage numbers in the follower and friend networks imply that there may not be much conversation happening between the governors and citizens with most governors not seeing the updates from citizens on their accounts at all.

Table 7: Reciprocity in friends and followers

Name	*Reciprocal Followers Count*	*Reciprocal Followers Percent*	*Reciprocal Friends Count*	*Reciprocal Friends Percent*
Martin O'Malley, MD	4	0.06	58	7.04
Brian Sandoval, NV	1	0.04	72	3.68
Deval Patrick, MA	18	0.1	212	1.84
Bill Haslam, TN	171	2.42	138	4.66
John Kitzhaber, OR	14	0.48	73	3.32
Pat Quinn, IL	9	0.74	20	1.02
John Hickenlooper, CO	1	0.04	86	8.66
Andrew Cuomo, NY	0	0	38	26.57
Chris Gregoire, WA	0	0	17	70.83

3.4.2 Graphical homophily

We sought to determine the locations of the users for each governor as well as who are they communicating with. *Figure 4* provides detail information of first level followers for Governor O'Malley and which counties or cities those users are located. More than ninety-eight percent of O'Malley's followers are geographically located within the United States. It should be observed from *Figure 5* that a preponderance of Governor O'Malley's followers is located in the metropolitan cities of Baltimore and the capital, Annapolis. For example, the metropolitan cities of Baltimore and Annapolis with a population of 620,961 and 38,390 have 1,175 and 160 followers, respectively whereas Columbia and Gaithersburg with populations of 99,615 and 59,333 have 51 and 25 followers. Moreover, one can observe from Figure 4 **Figure 5**for Governor O'Malley that the outlying areas of Salisbury and Ocean City have very few followers. Similar results were obtained for Massachusetts (*see Figure 5*). We found that followers for most (with the exception of Maryland and Nevada) governors are predominantly within their states as well as from the large metropolitan cities of these states. The Maryland governor had a big following in Washington D.C. (which is closely linked to Maryland) and the Nevada governor had following in California (which is a more active state on social media than the governor's home state of Nevada).

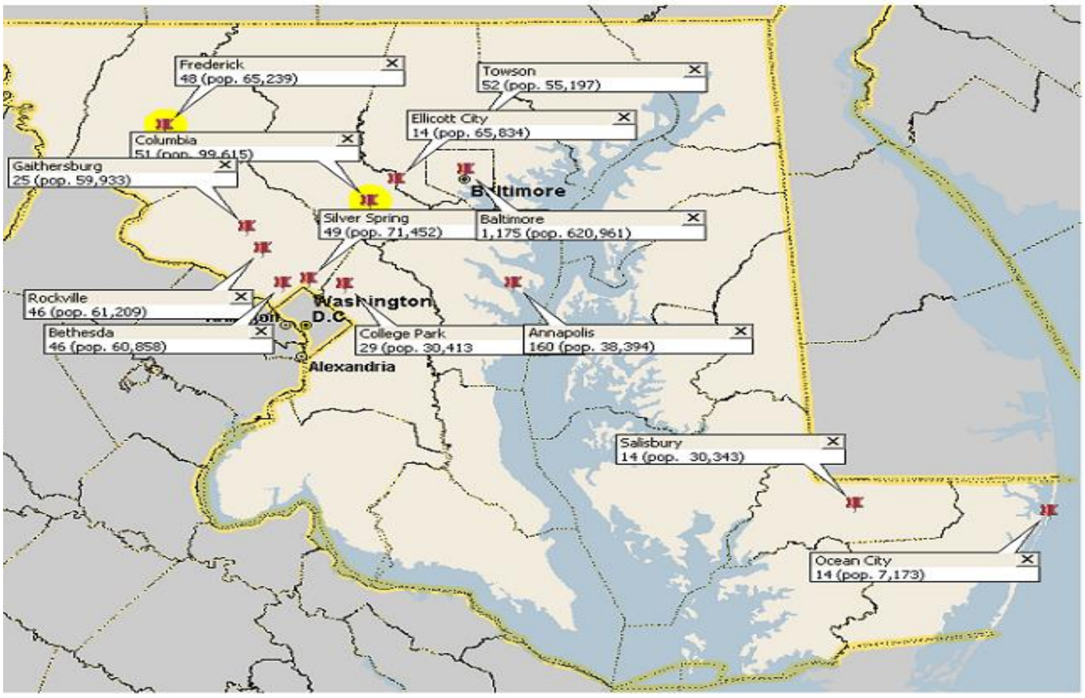

Figure 4: Location of Maryland governor's followers

Note: There were no followers from western Maryland so that part of the map is cropped out.

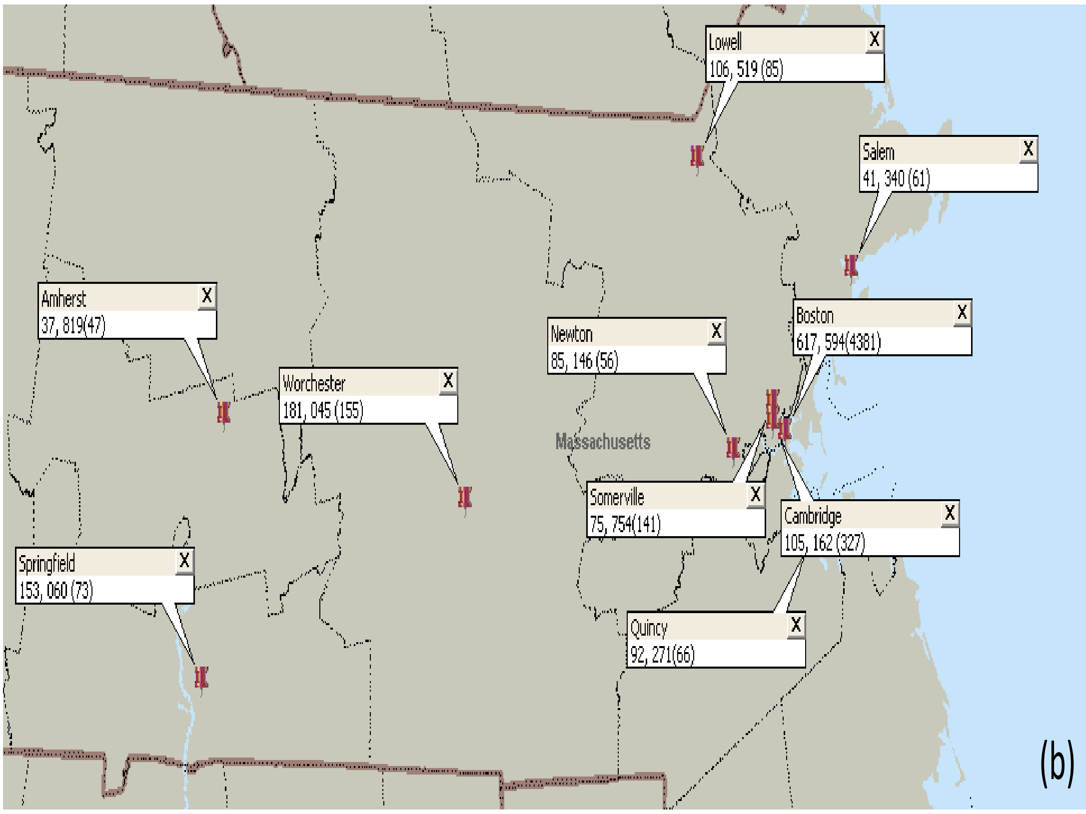

Figure 5: Location of Massachusetts's governors twitter followers

But who are the Governors following and where are these individuals located? *Figure 6* shows a basic analysis of the Government Agencies that are friends of the Governors. Our study revealed that a considerably large number of governors follow the National Aeronautics and Space Administration (NASA). Since NASA's last space shuttle mission was in the summer of 2011 as well as NASA's intense marketing campaigns targeting policy makers and school students to the Sciences along with its revamped website showing astonishing space photos and stunning live streaming videos of space expeditions and the capability to follow the expeditions interactively from earth using the social media such as Twitter and Facebook, may explain this phenomena.

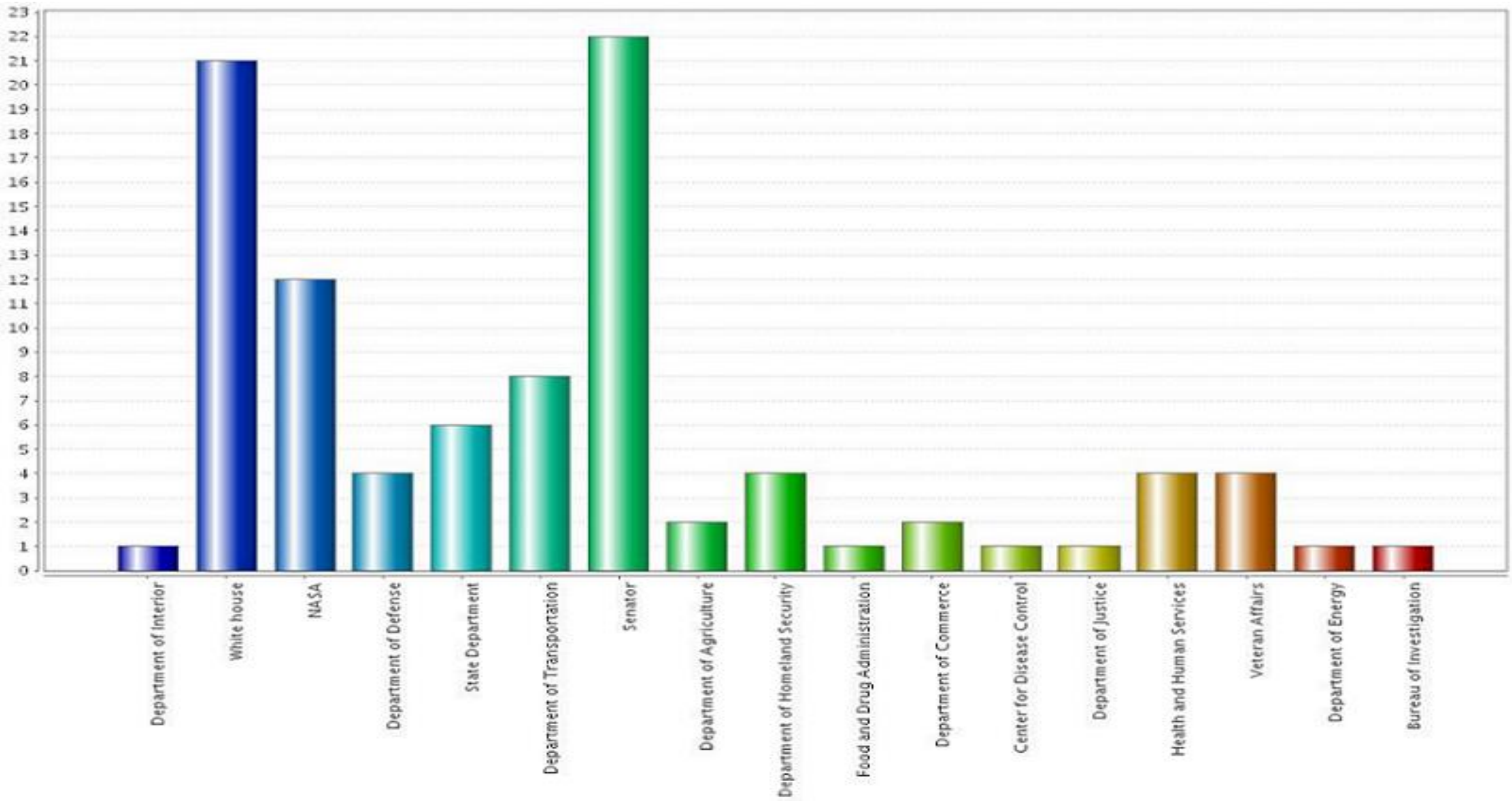

Figure 6: Agencies being followed by Governors

3.5 Conclusions and future directions

In the eight month period between March and early October, 2011, we downloaded more than seven million followers and four million friends for a total of over six million unique users for ten state governors. We observed that a certain level of homophily exists between the governors' followers and governors' friends as it relates to geographic location and political or social status. We also found that more than sixty percent of each governor's followers reside in their state. Further, over ninety-eight percent of all state governors' followers are from the United States of America. In addition, we discovered that the governors who are affiliated to their political party top brass have a large followers and friends base.

U.S. state governors have been utilizing Twitter and other latest technology in disseminating daily government information and government notices to its citizens on a regular basis. They appear to be reaching a certain portion of its population, because the majority of their followers are within their states. However, the fact that the governors are reaching the metropolitan cities at an appreciably higher rate than the smaller, rural communities who need the services most should be of some concern. We conclude that they have to do a better job in adding utility to their Twitter accounts so that citizens in any part of their state could use this medium to obtain state updates.

For future work, it would be worthwhile to know why state governors are following certain individuals and state agencies. Knowing why the governors are following the agencies could be beneficial to the federal government when disseminating human, economic and other valuable resources to its citizens.

CHAPTER FOUR: RESEARCH COLLABORATION IN CYBERSECURITY EDUCATION

PHASE I: EXPLORATION OF THE COLLABORATIONS IN THE COLLOQUIUM FOR INFORMATION SYSTEMS SECURITY EDUCATION

4.1 Introduction

Software security issues and vulnerabilities have been a challenge for business organizations, governments and educational institutions for decades. Much of the Information Assurance Education (IAE) and research is focused on college curriculum with a focus on program articulations, program assessments, and educational standards and guidelines [105]. There have been numerous publications on the development and the integration of information assurance in the computer science curriculum and other work addresses the success of these courses but also the extent to which our students are trained to handle security challenges [105]–[108].

Scholars and practitioners from educational institutions, government, and industry have collaborated within and across institutions and geographical boundaries on the above IAE topics. These collaborations among researchers represent a complex network structure with distinct structural characteristics. Social Network Analysis (SNA) methods can be used to study the levels of collaboration in the community, identify key institutions, individuals, and other stakeholders and study the growth of the information assurance education discipline. The focal point of this study is to analyze the research collaborations of IAE researchers who presented papers in Colloquium for Information Systems Security Education (CISSE) during the last decade (2001 to 2011). Given that CISSE has been at the forefront of IAE, we think it provides an ideal test bed to study the community using SNA methods.

We explore the following research questions: (1) what are the patterns of interactions between members of the IAE community? (2) are there core sets of active members who sustain the network?; and (3) how has the network evolved over time? The next section contains a review of social network analysis literature and methods.

4.1.1 Why study research collaborations?

Research collaborations by virtue of its inherent structure and the overall characteristics of its components, can be regarded as a constantly evolving network community since the members of any research network are free to enter and leave the community at any given time. Further, the relationship of key group members affects the dynamics of information flow and the propagation of information throughout the primary network and its sub-groups. Therefore, by studying the collaborations of researchers we are explicitly determining the characteristics of a network that is dynamic and evolving in nature.

In addition, social collaborative networks consist of participants or actors, depicted as nodes and the relationships between those actors, called links. By knowing the time which nodes and links that are added to the network structure at any given moment, is crucial to understanding the network dynamics of the collaborative research community.

Moreover, the process of conducting collaborative research is inherently social in nature. According to DeSanctis, the active and changing nature of a research community as well as its membership and activities represent its life [109]. Therefore, in order to evaluate the status and progress of the discipline, it is imperative to understand the social dynamics and the components of that research community over time [66], [109].

4.1.2 Information assurance research

Security issues and information technology vulnerabilities have been a challenge for organizations, governments and institutions for decades. Although the pervasiveness of information security vulnerabilities continues to persist at even a more alarming rate, much of the focal point of information assurance education and information assurance (IA) research have been centered on college curriculum and the integration of information assurance to existing computer science programs of colleges, program articulations, program assessments, and educational standards and guidelines. The National Colloquium for Information Systems Security Education (NCISSE) formed in 1996, since 2002 The Colloquium for Information Systems Security Education (CISSE) to reflect more international participation provides a forum for dialogue among leading security figures in government, industry, and academia. All research papers since 2001 are at http://www.cisse.info/archives.

In the United States the Department of Homeland Security (DHS) and the National Security Agency (NSA) have designated the National Center of Academic Excellence in Information Assurance Education (CAE/IAE) as the governing body for regulating the quality of information assurance of academic institutions [110]. The importance of information assurance and information assurance education is not only limited to the

US or US institutions but to also to other foreign countries such as the United Kingdom, Australia, New Zealand and Canada [110].

The emergence of IA had its roots way back since the onset of cryptography and computer telecommunications security. Then in the 1970s, faculty members from various US four-year colleges and universities began to develop and adopt security courses at their institutions. Later in the 1980s, research faculty from the universities began informal meetings to broaden their knowledge of the subject. It was during that time that research publications in the area of computer security begin to emerge [110].

Numerous publications have been made or written on what constitutes the course work of an IA program [106], [107], [110], [111]. The researchers have identified such courses as database management systems, networking and cryptography just to name a few. Other papers have examined the course objectives in establishing the guidelines necessary for effective program assessment, and establishment of virtual labs for testing purposes [106]–[108], [112].

4.1.3 Research questions

The focus of our study is to analyze the research collaborations of information assurance education (IAE) of researchers who coauthored Colloquium for Information Systems Security Education (CISSE) papers. Specifically, we examine the structure of their associated networks to determine the social identity of the discipline using social network analysis (SNA). The following are the primary research questions:

- What are the patterns of interactions between members of the CISSE community?

- Are there core sets of the “critical mass of active members who sustain the network”?
- How has the network evolved over time?

4.12 Literature review

See Section 3 of Literature Review Chapter

4.13 Research design

The figure below (*Figure 7*) illustrates the research design for analyzing the CISSE collaboration network for the proposed research questions. In each of the following subsections we provide a synopsis of how the data was acquired; the analysis of the network and; the patterns of structural interaction within the network.

4.13.1 Data acquisition

We designed a web crawler specifically to crawl the web and download the portable document files (PDF) to the hard drive was used to create the testbed for the study. Thereafter, we cleansed the data before creating the network analysis.

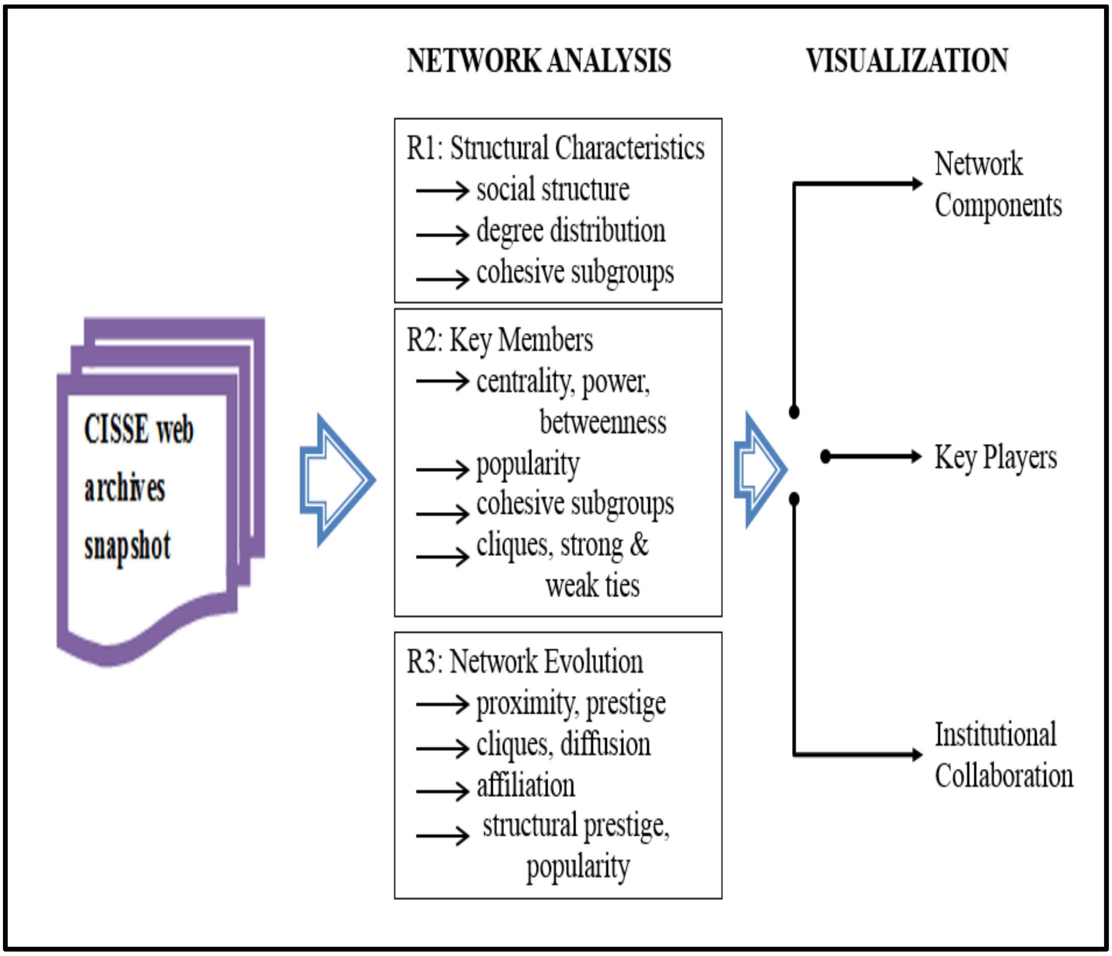

Figure 7: Research Design

4.13.1.1 Web crawler

The National Colloquium for Information Systems Security Education (NCISSE) was formed in 1996 to provide a forum for dialogue among leading figures in government, industry, and academia. Since 2002, to reflect more international participation, the organization changed its name to The Colloquium for Information Systems Security Education (CISSE). The Colloquium for Information Systems Security Education has an online archive of all it papers since 2001 (http://www.cisse.info/archives). The general information available to the public includes the paper, the title of the article, the author(s) of the article, the author(s) organization affiliation as well as the article abstract. We designed a web crawler and parser to download and store all the papers and associated meta-information for all CISSE papers published between 2001 and 2011 (except for 2003 where no papers were archived). This was done across several weeks so as to not burden the web server hosting the site.

All the PDF files for the period 2001 to 2011 were downloaded and saved to a computer hard drive. However, the specifics about those PDF files, including the name of the PDF file, the title of the paper, the paper abstract, the author(s) information and the organizational affiliation were downloaded into a database.

4.13.1.2 Data cleansing

Once the data was downloaded into the database, using structured query language (SQL), and other reporting tools such as Microsoft Excel and Access, we extracted the information that we considered to be pertinent to understanding and identifying the social structure of the CISSE social network. These include the name/s of author/s, their institutions of employment, the title of the publication, year of publication, and abstract of the published paper.

Table 8 presents a summary of the dataset. After performing name disambiguation based on institutional affiliation, data for a total of 348 unique authors from over 115 institutions was obtained. A total of 215 papers were published with 113 of them being collaborative with multiple authors.

Table 8: Summary of CISSE Dataset

Country	**Institutions**	**Authors**	**Papers**	**Collaborative Papers**
USA	125	334	198	122
Others	6	15	17	11

4.13.1.3 Network extraction

Using various popular social networking graphing and analysis software we examined the patterns of interactions between the authors, collaborating authors, as well as the respective institutions those authors represent. The tools used include UCINET, Pajek, R, NetDraw and KeyPlayer.

UCINET is a shareware software package used for the analysis of social data. The package is capable of handling both one-dimensional (1-mode) and two-dimensional matrix (2-mode) data. Although the software can handle a maximum of 32,767 nodes, the application can get extremely sluggish when it approaches ten thousand nodes. UCINET comes with a wealth of methods that can calculate centrality measures, the identification of subgroup, role analysis, permutation and other graph theory analysis.

Pajek is the closest rival to UCINET. It is a social network drawing freeware program that has many of the same capabilities as UCINET. However, there are several advantages to its rival. First of all, Pajek is made more specifically to handle larger networks than UCINET. Further, Pajek comprise of macros that can be used to record repetitive tasks. Finally, data can be sent directly to R to calculate additional statistics. R is a software package for statistical computation and graphics. It consists of a language plus a run-time environment with graphics, a debugger, access to certain system functions, and the ability to run programs stored in script files. One of the most common packages found in R is igraph which allows the visualization of network analysis.

NetDraw is a freeware tool for visualizing social network data. It can read both Pajek and UCINET files. NetDraw can save data to Pajek and to Mage (a graphical application). It can also save diagrams as EMF, WMF, BMP and JPG files. Printouts can be made directly from the program at high resolution which is of better quality than from printing document containing embedded graphics.

KeyPlayer is a freeware application for identifying an optimal set of nodes in a network. The two basic reasons for trying to find an optimal set of key nodes in a network are: (a) to identify targets for deletion, with the hope of crippling the network, and (b) selecting which nodes to either keep under surveillance or to try to influence via some kind of intervention. This is because well-connected nodes who are likely to possess a great deal of information, and who, because of their connections are in a position to influence others. KeyPlayer is distributed with Mage and Pajek.

4.13.2 Network analysis

The co-authorship network from the data was extracted with each node being an author and edge signifying a co-authorship relationship. The primary social network analysis concepts represented include: structural characteristics; key members and; network evolution.

4.13.2.1 Network growth and degree distribution

We examined the social structure of the network to analyze the change in the network as it grows and shrinks. The change in the number of authors, papers, and collaboration was plotted to study the change in the CISSE community. We used various social networking analysis and graphing software including UCINET [113], Pajek [5], R [114], [115], NetDraw [116], [117] and KeyPlayer [1] to calculate metrics and visualize the network.

We also analyzed the degree distribution of the network to determine whether it fits the power-law. A small-world network is characterized by a small average distance and a high degree of clustering coefficient. At the early stages of the CISSE collaborative network, we expect the network to display the small-world characteristics because researchers would initially collaborate with others they already know. Another degree distribution of the network could be represented by a power-law distribution that is highly skewed toward small degrees. We refer to this phenomenon as a scale-free network[14]. Many large social networks fall into this category. However, at this early stage of the CISSE network, we are unlikely to discover such trend.

A. Social structure

Social structure is considered to be a continuing pattern of ties among actors [3], [61]. The location of an actor in the social structure impacts his observed behavior and beliefs [3]. We employ a diversity of methods to establish the existence of a network. First off, we found the total number of papers that were published throughout the period. We then wrote a script to determine papers multiple authors. Using a grid, and SQL scripts we were able to discover authors who collaborated on multiple papers. Both UCINET and Pajek are excellent software alternatives for finding the social structure of a network. UCINET has a utility that enables the user to import an Excel or text file into the system for conversion to nine different data formats. After the data files were generated, we used the network tools functionality to ascertain the giant component of the network and the distances of the paths between the nodes.

B. Degree distribution

We also analyzed the degree distribution of the network to determine whether it fits the power-law. A small-world network is characterized by a small average distance and a high degree of clustering coefficient. At the early stages of the CISSE collaborative network, we expect the network to display the small-world characteristics because researchers would initially collaborate with others they already know. Another degree distribution of the network could be represented by a power-law distribution that is highly skewed toward small degrees. We refer to this phenomenon as a scale-free network[14]. Many large social networks fall into this category. However, at this early stage of the CISSE network, we are unlikely to discover such trend.

C. Cliques subgroups

Since the researchers in the network community typically share a common interest, exclusively security, we used cluster analysis to determine the extent those researchers form subgroups in the network. To find whether the network clusters are true subgroups, we analyze the collaboration using a network analysis statistical tool called cluster coefficient. With random networks it is highly unlikely to obtain subgroups within the network because each node has the same probability to connect to other nodes. Therefore, the cluster coefficient for random network is characteristically rather small compared to a scale-free or small world network.

Cohesion is a fundamental concept in understanding and explaining the underpinning of social network analysis. In fact, cohesion and the measurement of cohesive subgroups are the ultimate objective when attempting to analyze subgroups in networks. Several approaches may be used for identifying cohesive subgroups. A common method used to cluster actors is by focusing on the pattern of ties among the actors[61]. In the analysis of the cohesive subgroups of the CISSE network, we have decided to examine the ties within dyads and the roles actors in the dyads have on the larger network using a criteria by Prell[4]:

- The number of direct (indirect) ties linking individuals in the group together
- The relative isolation of groups to outsiders
- The extent to which individuals can reach each other
- The extent to which the group is vulnerable to fragmentation

The clustering coefficient of the network was found using both UCINET and Pajek. We also used Pajek to establish the cohesive subgroups of the network. To determine this relationship we must first remove the lines that form loops in the network. A loop refers to a line (a link between actors) that connects a vertex (an actor) to itself. We have decided to remove the loops because researchers who select themselves in the network do not form a true cohesive group in the social network setting. Then we convert the results into many useful social analysis diagrams such as the minimum spanning tree (MST) using igraph.

4.3.2.2 Key members

One of the assumptions of social network analysis is that the ties between actors can transfer various types of resources such as information and trust [61]. The implication of the strong or weak ties in social network analysis indicate that some key members in the network have the potential to keep the network together; may have the power in the diffusion of information throughout the system as well as the ability control subgroups within main network.

A. Central measures

We used two measures to determine the prominence of an author in the CISSE network: centrality and betweenness. Centrality scores for an individual in the network ranges from 0 – 1, with lower scores indicating lower levels of centrality. Betweenness is a measure of an actor’s position between pairs of vertices. It is a link in the chain of contacts that facilitate the spread of information through the network [5]. The more central an actor in the network, the more important he or she becomes to the flow of information in the communication network.

B. Popularity

Popularity is a measure of structural prestige in the network. It can also be said to be the number of choices a vertex receives. In other words, the number of its indegree directed at it. To determine the key players in the network, we loaded dataset into the KeyPlayer program. Then we start crippling the network by removing those key nodes. Once this has been done, we exported the network into Pajek for graphical display.

4.3.2.3 Network evolution

Using the historical dataset for the last ten years, we constructed snapshots of the network at various spatial intervals to determine how the network has evolved over time. Not only do we focus on the changing structure of the network, we also looked at the ability of the network to absorb new members because the organization needs to be able to attract new members if it is to continue to thrive for another ten or more years.

A. Proximity, prestige

Proximity is one of the commonly used measures of social network analysis. It is an index that reflects the distance of an actor from all vertices in its input domain of the network [5], [118]. Proximity significantly impacts groups and subgroup power, information diffusion and other group dynamics of a social network. The effect and the impact of direct and indirect neighbors are taken into consideration when examining proximity. The closer the individuals are to the input domain (meaning the shorter the path), the higher the prestige of that individual. We describe the concept of proximity prestige as a weight of each path distance to the vertex. Hence, a higher distance yields a lower prestige [5].

Highly productive researchers by virtue of their productivity are likely to develop some degree of prestige within the social network structure. In the same token, individuals who are close to those highly productive individuals are likely to publish research papers with those persons. The same applies for institutions. Therefore, the most productive colleges and other institutions are likely to have more colleges joining them in publication.

To calculate whether the network has been able to absorb new researchers each year, we analyzed the data for researchers who first published papers each year for ten years from 2001 and 2011. Then we created a trend analysis of the data to show a comparative picture of all three variables; namely new authors; total number of authors and total number of papers

B. Cliques, diffusion

Informal groupings of vertices within the subgroups are referred to as cliques. Cliques are extraordinarily interesting not only because members of the subgroup tend to develop strikingly cohesive structures but because they also develop their own form of norm, rules and culture that could be different from the larger network [4].

Another important aspect in social ties is how information flows through a social network structure. Social ties enable participants to gain access to information which they would not have otherwise obtained. Research has shown that people in crucial positions in the information network may strategically spread or inhibit the spread of information — information diffusion [5], [6].

To determine whether the structure of the network has changed with time, it is essential that we examine the cohesive subgroups within the network. Since the data is for a fairly short period (ten years), we have decided to divide it into three main temporal boundaries, namely *pre-2005, to 2008* and, *to 2011*. Data for those three time frames were used to construct a network diagram as well as the minimum spanning tree.

C. Affiliation

People and institutions often gather because of common interests or ties by association. This kind of social ties is referred to as affiliations. Affiliations could be structural relations that are generally forced by the various circumstances. They result more from private choices and less through friendship. When people gather around one or more organizations and events we generally refer to this as social ties [5]. McPherson et. al argued that as networks evolve over time through cumulative processes of tie creation and dissolution, multiple affiliations inevitable embed itself into the system [59].

We established the most productive publishers in the network by calculating the authors who have published the most papers. The same applies for institutions. The productive institutions are the institutions its employees published the most papers.

D. Structural prestige and popularity

When we inspect the nodes within the network, we observe nodes (authors) in which there are both input degree and out-degree of relations to other nodes in the network. Those nodes with the highest number of input degree relations are considered to be the most prestigious in the network. Prestige is often associated with power. One of the simplest measures of structural prestige is popularity. Popularity is a measure of the number of choices a vertex receives [5], [119]. Therefore, the higher the indegree of the nodes the higher the structural prestige.

Structural prestige and popularity have some unique characteristics that can help maintain or change the state of the social network. For example, because other actors tend to gravitate towards others because of their prestigious position, they help keep the social network intact. Conversely, the rivalry among actors could develop into social circles of cohesive subgroups that re-energize the structure of the network.

The degree to which authors collaborate with other authors, whether inter-institutional or cross-institutional, in publishing papers each year provides an insight into how the structure of the network has changed over the years. It also provides data that we can use to form a trend analysis of the stability and future growth potential of the network. Using the dataset a line graph was used to display the trend of a contribution of each value over time.

4.13.3 Visualization

Social network metrics such as density, centrality measures, clustering algorithms, and so on, are essential in analyzing a social network's dynamics [61]. Besides metrics, there are other tools that could be used for analyzing social networks. One such tool is network visualization. Network visualization can help us see and understand patterns that may not be readily apparent by simply looking at metrics [120]. Moreover, it helps us to communicate our findings more effectively [121]. Many of the social network analysis software have modules for network visualization. The visual representation of social networks is indispensable to understanding network data and to convey the result of the analysis [122]. Everton argue that network visualization algorithms are capable of identifying cohesive subgroups even when clustering algorithms are incapable of doing so [61]. Hence the reason why both metrics and visualization are complimentary parts of the social network analysis toolbox [61].

As indicated above, network visualization software is often packaged with other social network analysis software tools. In addition, network visualization may also be distributed as a standalone data analysis method. Although there is a vast array of social network analysis tools, we narrowed our usage to four: Pajek; UCINET; R and KeyPlayer; respectively. Pajek[1] is a freeware that is widely used for drawing large social networks. It has significant analytical capabilities and can be used to calculate most social network metrics. Pajek data can be sent directly to R for further statistics. R[2] contains several packages relevant for social network analysis. Igraph is one of its generic network analysis packages. We used igraph throughout our study specifically for drawing minimum spanning trees.

---

[1] Pajek can downloaded for free for noncommercial use from the Pajek web site: http://pajek.imfm.si/doku.php

[2] R is a versatile, open-source, SNA tool. It can be obtained from: http://www.r-project.org/

UCINET [3] is a shareware social network analysis tool. It is a comprehensive package for the social analysis of social network data as well as other 1-node and 2-node data. It is distributed with a visualization tool called NetDraw. NetDraw is an application developed by one of the creators (Steve Borgatti) of UCINET.

Finally, KeyPlayer[4] is a free application for identifying an optimal set of nodes in a network for two basic purposes. Firstly, it allows one to cripple a network by removing key nodes. Secondly, it also allows a user to select which nodes to either keep under surveillance or to try to influence via some kind of intervention.

A. Network components

To display the network components, we used UCINET and Pajek as our visualization tool. In addition, the cumulative frequency distribution was calculated using the igraph software package in the R network analysis software. The minimum spanning trees created in this section (and other sections) is the byproduct of the R social analytic package.

B. Key players

The key players in the distribution were determined through various ways including the removal of some of the key players from the network subgroups using the KeyPlayer application.

---

[3] UCINET is distributed by Analytic Technologies at http://www.analytictech.com/

[4] KeyPlayer is freely distributed. Its author is Steve Borgatti. Latest versions of the program may be found at www.analytictech.com.

C. Institutional collaboration

Some of the institutional and cross-institutional collaboration analysis was done using all of the social analysis software mentioned above.

4.14 Results and discussion

4.14.1 Structural characteristics

Our results revealed the CISSE collaboration network as a structurally sound and well defined network. The network collaboration contained a giant component accounting for 95.43 percent of all the authors who have published articles in CISSE. Thirty-one percent of the authors collaborated on projects within groups comprising of 2 to 9 members. Only 6 percent of the CISSE authors have never collaborated with anyone except themselves. There were multiple collaborations among the same authors. In fact, collaborations between two authors occurred fifteen times with the highest number of collaborations between any pair being five. The average number of researchers working on an article is three. See *Figure 8* below.

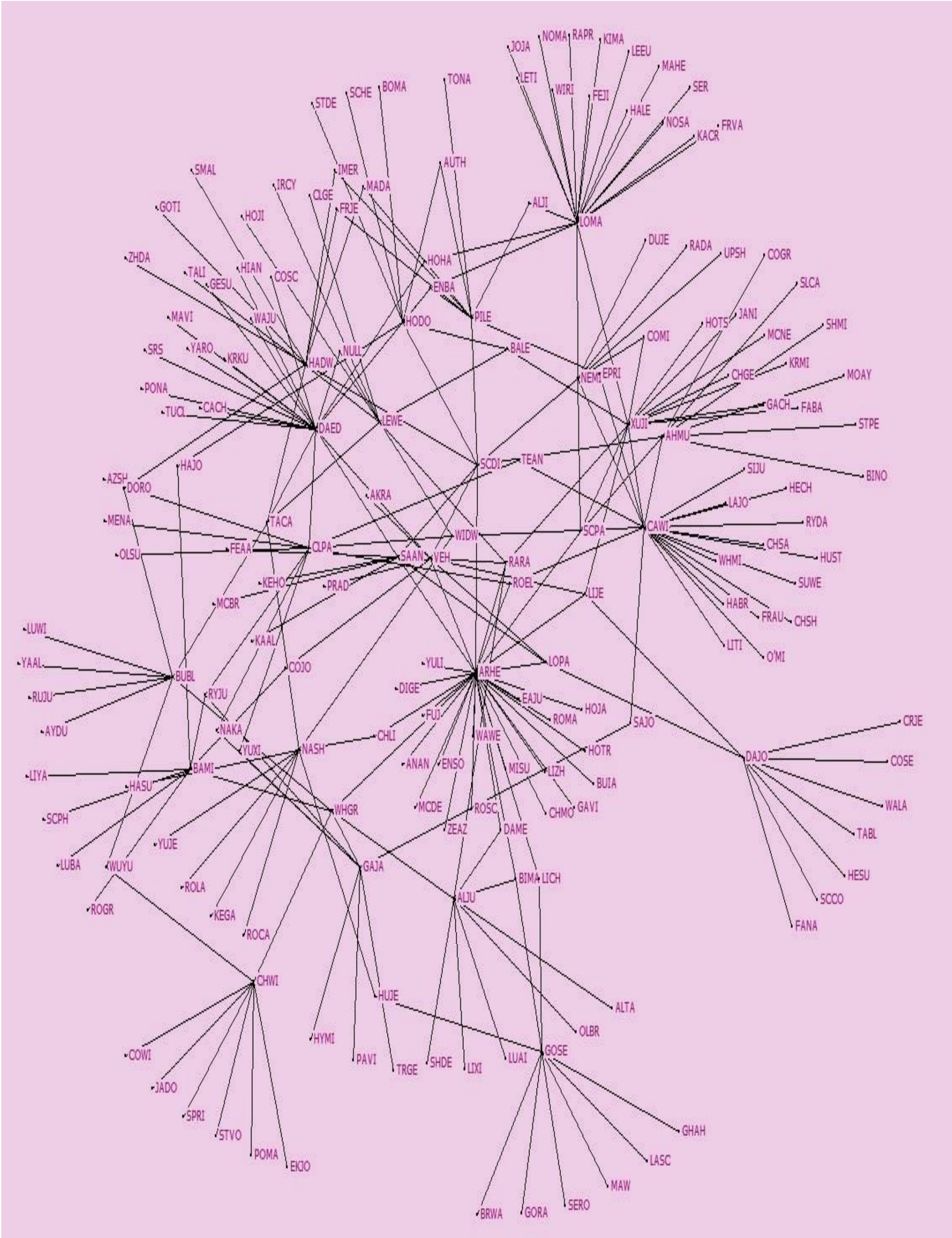

Figure 8: Network Components of Authors

Two of the measures, *distance* and *efficiency*, can be used to determine information diffusion through the network. Whereas distance is defined by the shortest path between two nodes [123], the efficiency of a network is the sum of the inverses of lengths of all-pairs shortest paths [124]. We found that the average distance in the giant component was 3.34 while the closeness or efficiency was 0.42. In order to compare the random network with its counterpart, we generated thirty random networks of the same size and average degree with those of the giant component. We found a resulting mean distance of 1.1 and an efficiency of 0.95. Essentially, the researchers in the CISSE community and the random network are almost identical. However, the efficiency of the CISSE community is significantly lower than the random network.

RQ1.2: What structural characteristics does the collaboration network have?

Networks structures vary depending on their type of collaboration. For instance, in a random collaboration, researchers select collaborators on a random basis. On the other hand, with a scale-free structure, a few active members disproportionately dominate the activity within the network. We anticipate the CISSE community to be of a scale-free network structure.

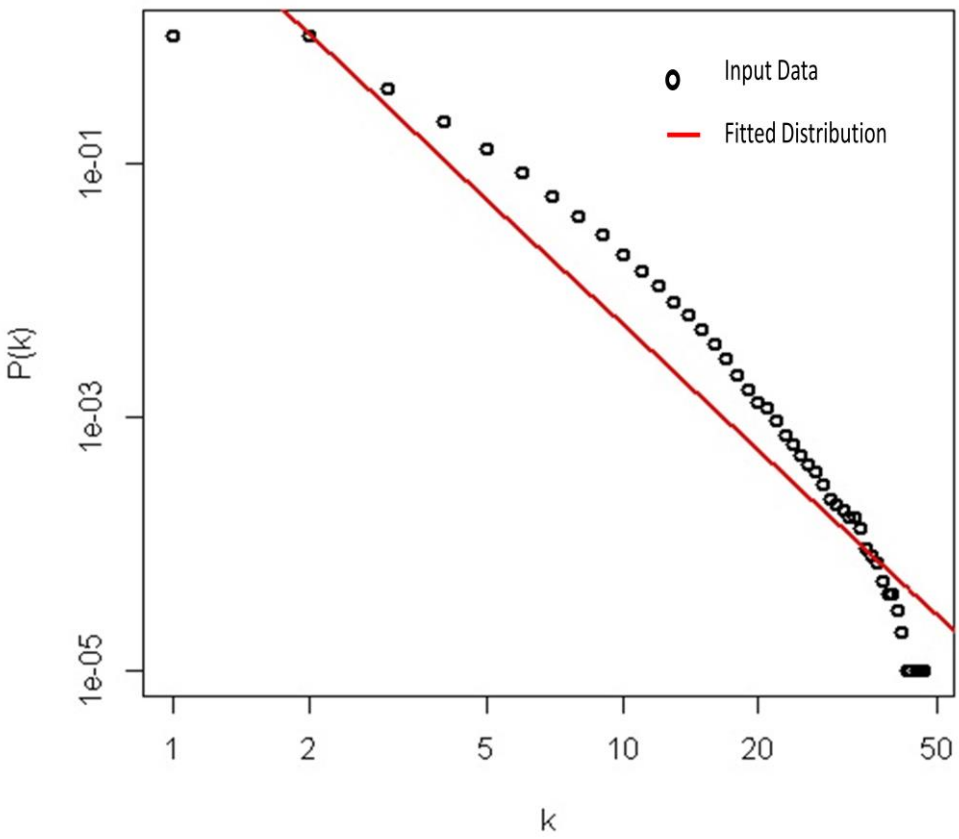

Figure 9: Cumulative Degree Distribution of CISSE Collaboration Network

Figure 9 illustrates the cumulative degree distribution of the CISSE collaboration network. As can be observed, the diagram clearly represents a distinctive scale-free characteristic with a few authors having a large number of collaborations and many authors with a significantly small number of collaborations. *Figure 9* is evident that the collaboration network is not a true power-law relationship because the tail shows a steep decline in the number collaborators. The results show that the CISSE network is not structurally cohesive because collaboration between members are disproportionate in size and structure. Further, the findings indicate that researchers have not been able to attract other researchers from other institutions and countries on a consistent basis.

RQ1.3: Are there subgroups in this community?

Since the researchers in the CISSE community typically share a common interest, exclusively security, we can use cluster analysis to determine the extent those researchers form subgroups in the network. To find whether the network clusters are true subgroups, we analyze the collaboration using a network analysis statistical tool called cluster coefficient. With random networks it is highly unlikely to obtain subgroups within the network because each node has the same probability to connect to other nodes. Therefore, the cluster coefficient for random network is characteristically rather small. The cluster coefficient for the CISSE network was 0.0077 compared to the 0.082 of its random subgroup. Essentially, this means that very little or no cliques have been formed among researchers from different nodes. This could be attributable to the fact that the organization has not been in existence for a very long period of time and thus researchers have not been able to form strong cliques within the CISSE community. Our findings indicate that the "small world" phenomenon is less conclusive in the CISSE community although, in general, the overall short distance for the subgroups indicates that there is a greater evidence of a "small world" than for the CISSE community.

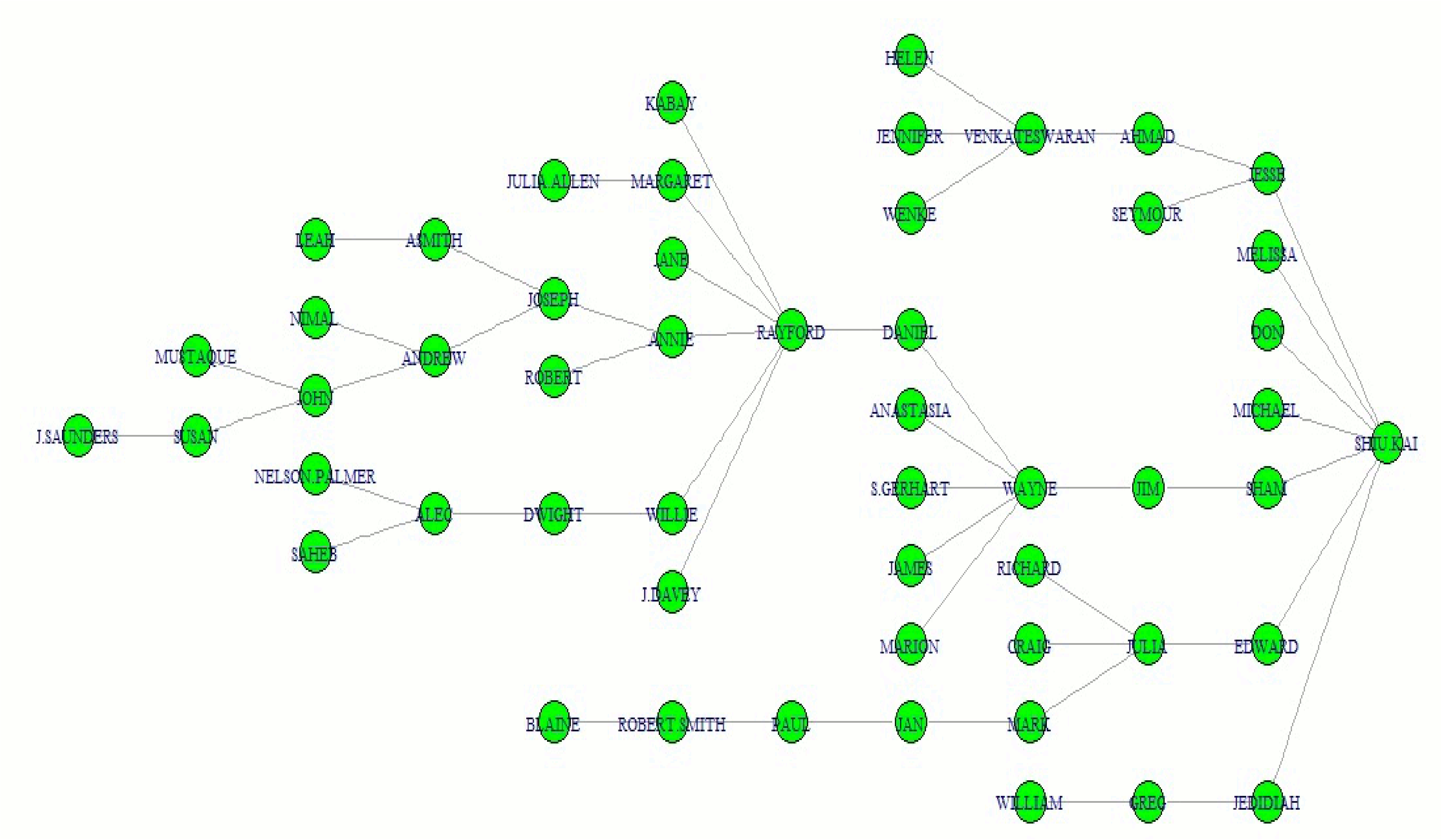

Figure 10: Minimum Spanning Tree of CSSE Collaboration Network

RQ2: Are there core sets of the "critical mass of active members who sustain the network"

RQ2.1: Who are the most eminent researchers in the community?

Eminent researchers are those researchers who may have published several papers either on their own or perhaps with many collaborators. We have decided to analyze an author's productivity based on three factors used by Xu and Chau, namely; *normal*, *straight* and *adjusted ranks*. Normal rank is determined based on the number of papers an author publishes. On the other hand, straight rank counts only the papers of which the author is the first author. Adjusted ranks is built on collaborative strength of which the author receives 1 for a paper with only two authors but a less than 1 value with papers with more than two authors [72]. *Table 9* below lists the top authors in each productivity category. The combination of all three productivity ranking should give an unambiguous representation of who the most productive researchers are in the CISSE community. The results denote a distribution of author productivity is a power-law with an absolutely meager R-squared of 0.0064. In essence, this means that there is virtually no correlation between authors and the numbers of papers published. Thus, our findings prove that just a very few authors have published a large number of papers.

Table 9: Author Productivity Rank

Name	# Papers	Name	# 1st Authored Papers	Name	Adjusted Score
1. Dino Schweitzer	9	1. Richard Epstein	5	1. Gregory White	6.5
2. Gregory White	8	1. Paul Schembari	5	2. Dino Schweitzer	6.08
3. Kara Nance	7	2. Patricia Logan	3	3. Kara Nance	5.5
3. Richard Epstein	7	3. Alec Yasinsac	2	4. Helen Armstrong	5.0
4. Brian Hay	6	3. Jill Slay	2	5. Brian Hay	4.5
4. Helen Armstrong	6	4. Matt Bishop	1	6. Matt Bishop	3.0
5. Patricia Logan	5	4. Helen Armstrong	1	7. Carol Taylor	2.75
5. William Caelli	5	4. Jane Jorgensen	1	8. William Caelli	2.5
6. Carol Taylor	4	4. Matt Bishop	1	9. Richard Epstein	1.5
6. Matt Bishop	4	4. Carol Sledge	1	10. Mike O'Leary	1.18

Table 10: Author Centrality

Degree Centrality	# of Collaborators	Closeness	Betweenness
1. Dino Schweitzer	9	1. Gregory White	1. Gregory White
2.Gregory White	7	2. Helen Armstrong	2. Helen Armstrong
3. Kara Nance	6	3. Brian Hay	3. Richard Epstein
4. Brian Hay	6	3. Kara Nance	4. Brian Hay
5. Helen Armstrong	6	4. Richard Epstein	4. Kara Nance
6. William Caelli	5	5. Matt Bishop	5. Jill Slay
6. Mike O'Leary	4	6. Nimal Jayaratna	6. Dino Schweitzer
6. Blair Taylor	4	6. John Hamilton	6. Jeffrey Livermore
7. Carol Taylor	4	7. Patricia Logan	7. Carol Taylor
7. Matt Bishop	4	8. Ronald Dodge	8. Melissa Dark

To determine the prominence of a researcher, we have decided to use degree centrality of the CISSE network. The degree of a researcher is the number of collaborators he or she has. In addition, we utilized two other commonly used centrality measures: *closeness* and *betweenness*. Both of these measure the reachability of a person within a network. Betweenness, measures the extent to which one node lies between other nodes in the network. For instance, how significant a person is to the transmission of information through a network as well as the extent to which a person may control the flow of information due to his or her position in the communication network can be traced back to his or her betweenness. Closeness, on the other hand, measures the distance from one node to all other vertices within the network. If you want to measure indirect contact to neighbors in the entire network, then closeness is the preferred measure. The significance of the vertex in the network analysis is best captured using betweenness centrality. *Table 10* above shows a considerable number of researchers are high in all three centrality measures, undoubtedly proving their structural roles.

RQ2.2: Are these star researchers critical to holding the community together?

Unlike a structural cohesive environment in which everyone plays an equally important role in the network community, actors in a scale-free network have clearly defined role which enable network connectivity. To determine the central roles of key researchers in the network, we decided to perform a *"network robustness test"* [17] to observe how the network would alter if those main researchers are missing. Fundamentally, our hope is to identify an optimal set of key nodes. If we progressively delete those key nodes from the network, at some point the network should completely cripple the network. Further, since well-connected nodes are likely to possess a great deal of information, and who, by virtue of their connections are in a position to influence others, breaking the connections would ultimately cause members to have a difficult time to communicate. Moreover, removing key nodes would inevitably cause the network to become small clusters leading to an extremely small average node distance.

We first identified the key researchers in the network using KeyPlayer1[125] and presented the diagrams using NetDraw. The top ten key players in the CISSE network are presented in *Figure 11* below. We then removed the nodes for the key players in the network not only to increase the distance between some pairs of nodes but to completely disconnect them as well. *Figure 12* clearly shows the disconnectedness and fragmentation of the CISSE network. Finally, the results of the removal of those key players left us with a non-cohesion measure of 0.973 percent. One should note here that the closer the non-cohesion measure is to 1 the greater the optimization and the smaller the number of clusters within that network.

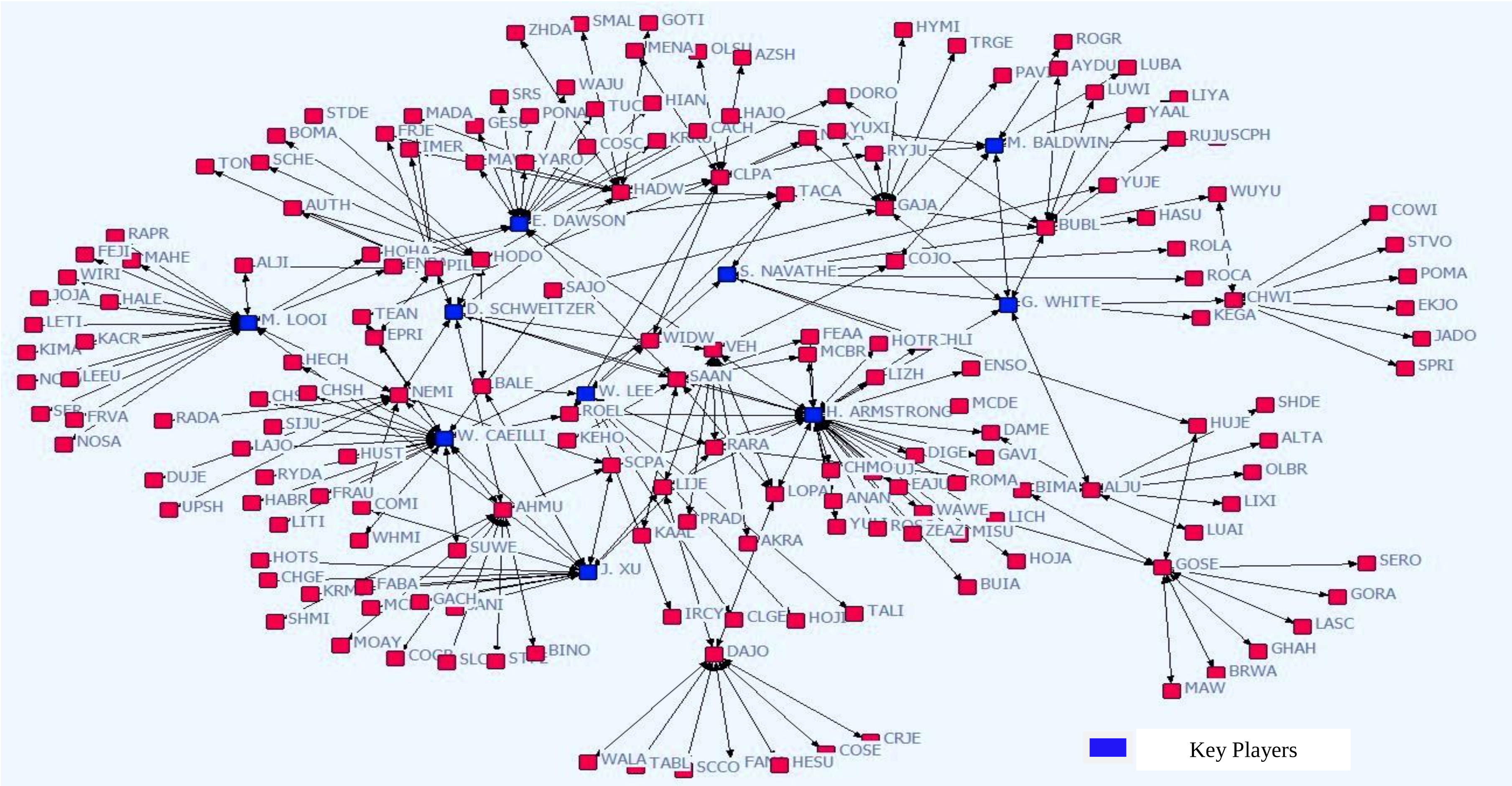

Figure 11: Key Players in CISSE Network

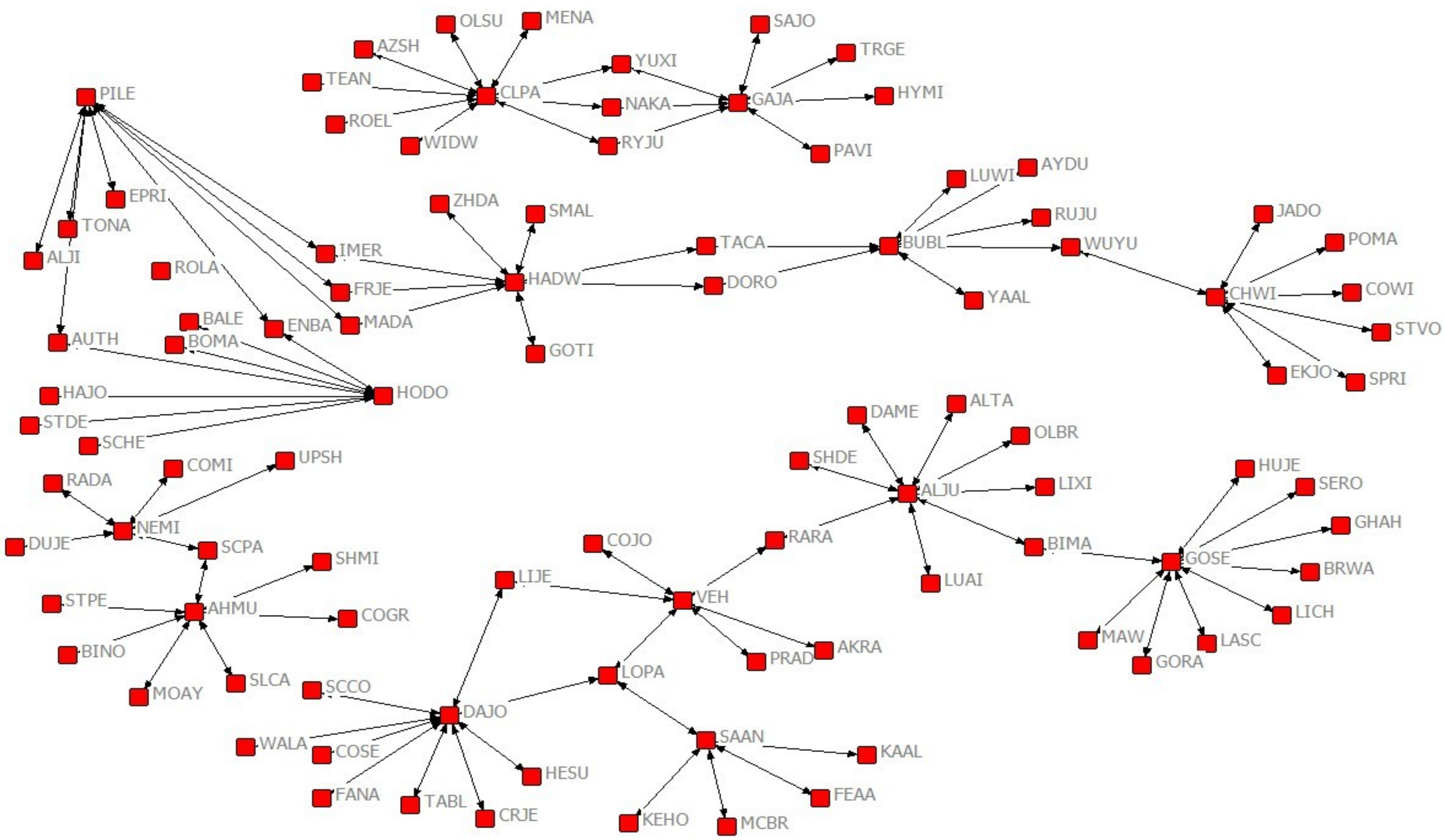

Figure 12: CISSE Network after Deleted Key Players

RQ2.3: Which institutions are the most productive?

Table 11: Top 10 Productive Institutions

<table>
<tr><th>Institution</th><th># of Papers</th><th>Institution</th><th># of First-Authored Papers</th></tr>
<tr><td>1. United States Air Force Academy</td><td>26</td><td>1. University of North Carolina at Charlotte</td><td>16</td></tr>
<tr><td>2. Towson University</td><td>23</td><td>2. United States Air Force Academy</td><td>10</td></tr>
<tr><td>3. University of North Carolina at Charlotte</td><td>19</td><td>2. Naval Post Graduate School</td><td>10</td></tr>
<tr><td>4. North Carolina A&T State University</td><td>17</td><td>2. North Carolina A&T State University</td><td>10</td></tr>
<tr><td>5. The Pennsylvania State University</td><td>16</td><td>3. United States Military Academy</td><td>9</td></tr>
<tr><td>6. The University of Texas at San Antonio</td><td>15</td><td>3. The Pennsylvania States University</td><td>9</td></tr>
<tr><td>6. University of Alaska Fairbanks</td><td>15</td><td>4. Towson University</td><td>8</td></tr>
<tr><td>6. Naval Post Graduate School</td><td>15</td><td>4. Georgia Institute of Technology</td><td>8</td></tr>
<tr><td>7. United States Military Academy</td><td>13</td><td rowspan="2">5. Auburn University
5. Boston University
5. Rochester Institute of Technology</td><td rowspan="2">7</td></tr>
<tr><td>8. University of Idaho</td><td>11</td></tr>
</table>

We measured the universities' productivity using both normal and straight rank. *Table 11* above presents the top ten productive institutions. The United States Air Force Academy, the University of North Carolina at Charlotte, North Carolina A&T State University and The Pennsylvania State University are high in both categories. Other institutions in the top category include Towson University, the Naval Post Graduate School and United States Military Academy. The results from the study illustrate that none of the top ten institutions from either group is from another country other than the United State of America.

RQ2.4: What are the cross-institutional collaboration patterns?

The overall existence of CISSE was to provide a forum for dialogue for leading figures in academia, government and industry in the area of information security education. Researchers, who ultimately have that common interest, should therefore find it relatively straightforward to collaborate not only within their institutions but on an international scale as well.

An analysis of the researchers who collaborated with other researchers that were not a faculty member of their institutions was examined. Using that data, a network diagram () and a minimum spanning tree (MST) () were constructed on all researchers who participated in cross-institutional security papers. An institution ***A*** is considered to be connected to institution ***B*** if authors from both institutions have collaborated in at least one paper. The network collaboration links are based on the existence of an alliance in publishing rather than the frequency of their publication.

The giant component for the collaboration network comprise of 78 institutions which accounts for 60 percent of the institutions that have published in CISSE. There are 212 links among these institutions. On average, an institution collaborates with 1.64 other institutions. The top two collaborative institutions are the University of Idaho and Brigham Young University, which have collaborated with 9 and 8 other institutions, respectively.

In *Figure 13* and *Figure 14* below, the collaborative network and MST for the institutions are presented. The nine most prolific institutions are labeled with their names. It can be observed from *Figure 13* that the majority of the institutions that published research paper together form a concentrated web and are at the heart of the collaboration network while a few institutions including Arizona State University and the University of Memphis form a small outlier group.

The minimum spanning tree (MST) for the cross-institutional collaborative network for CISSE researchers illustrates some very remarkable results. First of all, there is an incredibly low level of research collaboration between global institutions. In fact, the lone alliance is between the United States Military Academy, Curtin University of Western Australia, Manchester Metropolitan University and Edith Cowan University of Australia. Further, cooperation between institutions within the United States is not limited to geographic proximity. For example, Eastern Washington University teamed up with Rochester Institute of Technology (New York), Southern Polytechnic State University (Georgia) and the University of Idaho. In addition, a close examination of the MST will show that although three of the top nine collaborative institutions, United States Military Academy, Pace University and Rochester Institute of Technology, are located in New York, they are in the extreme ends of the network.

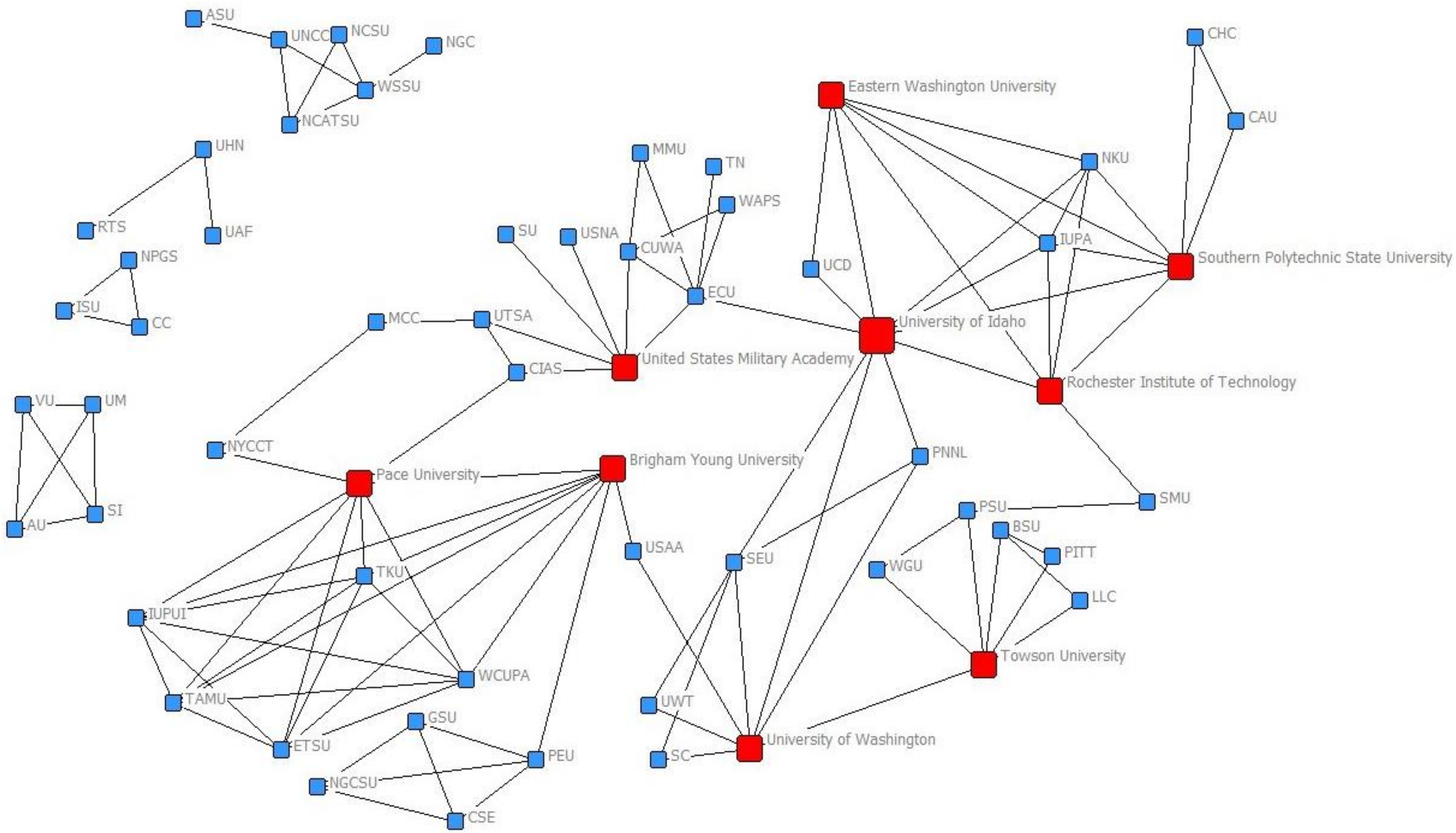

Figure 13: Cross-institutional Collaboration Network

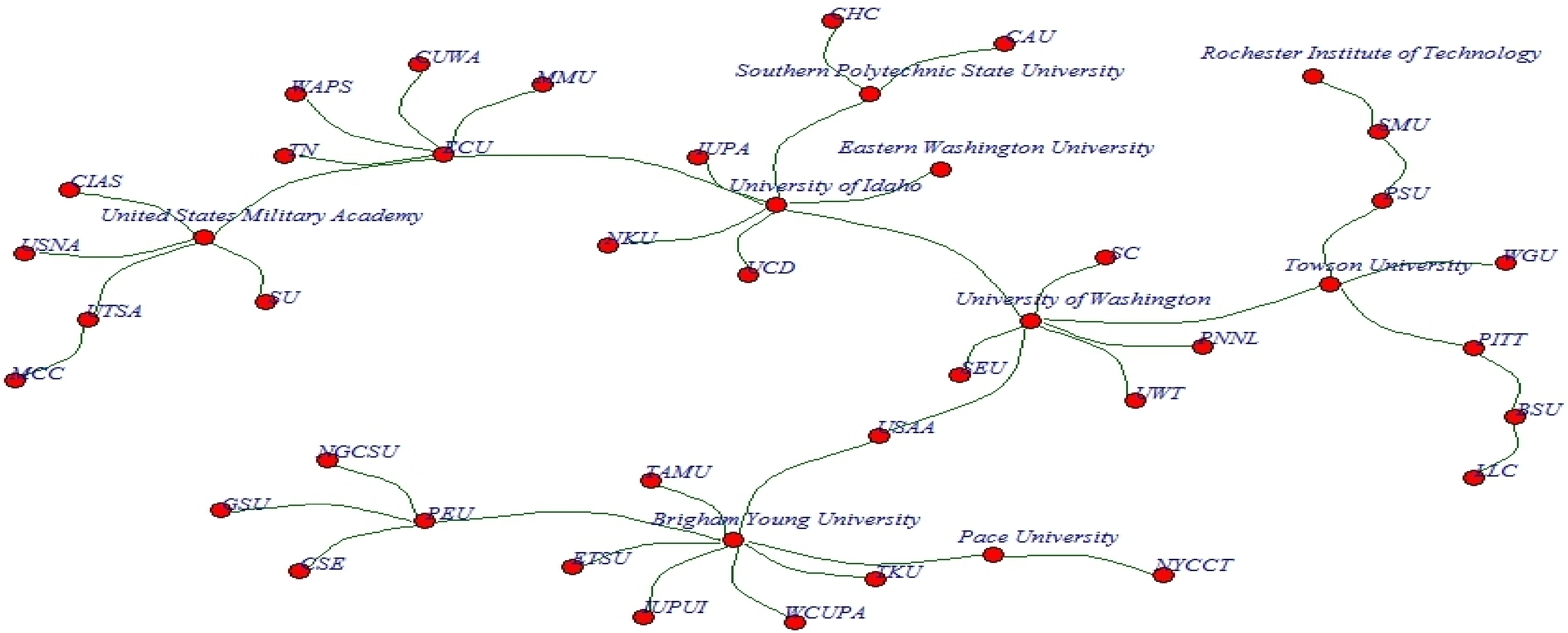

Figure 14: Cross-institutional Collaboration Network (MST)

WCUPA - West Chester University of Pennsylvania, PSU - Pennsylvania State University, SMU - Southern Methodist University, SEU - Seattle University, CUWA - Curtain University Western Australia, USNA - United States Naval Academy, PEU - Purdue University, USAA - University of Southern Australia

RQ3: How has the network evolved?

RQ3.1: Has the community been able to absorb new members?

The ability of the CISSE community to attract novice researchers, whether doctoral students, new faculty members, or researchers who collaborate with other researchers on IS security, is a sign of an emergent CISSE research community. *Figure 15* below presents a pictorial analysis of the number of authors who publish in CISSE, the number of new authors who publish in CISSE for the very first time, and the number of papers accepted or published each year. Overall, the plotted diagram gives an indication of the trend analysis of the CISSE for the ten year period, 2001 to 2011. The average number of papers accepted for publishing for that period is about 22. The graph for *Figure 15* illustrates a steady increase in the number of papers published between 2002 and 2006. This was followed by a drop of almost twenty-seven percent in 2007. Thereafter, we see another steady increase in publishing until it peaked at 32 for 2010. During 2001 and 2011 both the number of authors and new authors are directly proportional to the number of papers published. Hence, an increase in the number of new authors results in an increase in the number of papers accepted. The inverse is also true, when the number of papers accepted declines; we also see a reduction in the number of new authors.

The graph below (*Figure 15*) also shows that throughout the first five years of the program, the number of new authors and the total number of authors were almost neck and neck. This near identical number explains the fact that researchers may have been using that time frame to experiment the program. Once that investigational phase has been passed, the researchers are thus able to determine where possible collaborative efforts may exist.

Overall, it is difficult to pin point the factors that would lead to the dramatic decline in the number of new authors and other collaborating authors between 2006 and 2007 or the sustained increase in the number of authors from 2007 and 2010. Unless the committee is placing a cap on the number of papers received, we believe that neither location nor the dates for the colloquial proceedings played any instrumental role in the number of papers published since all colloquiums have been held in the United State while the dates are always set for annually in June.

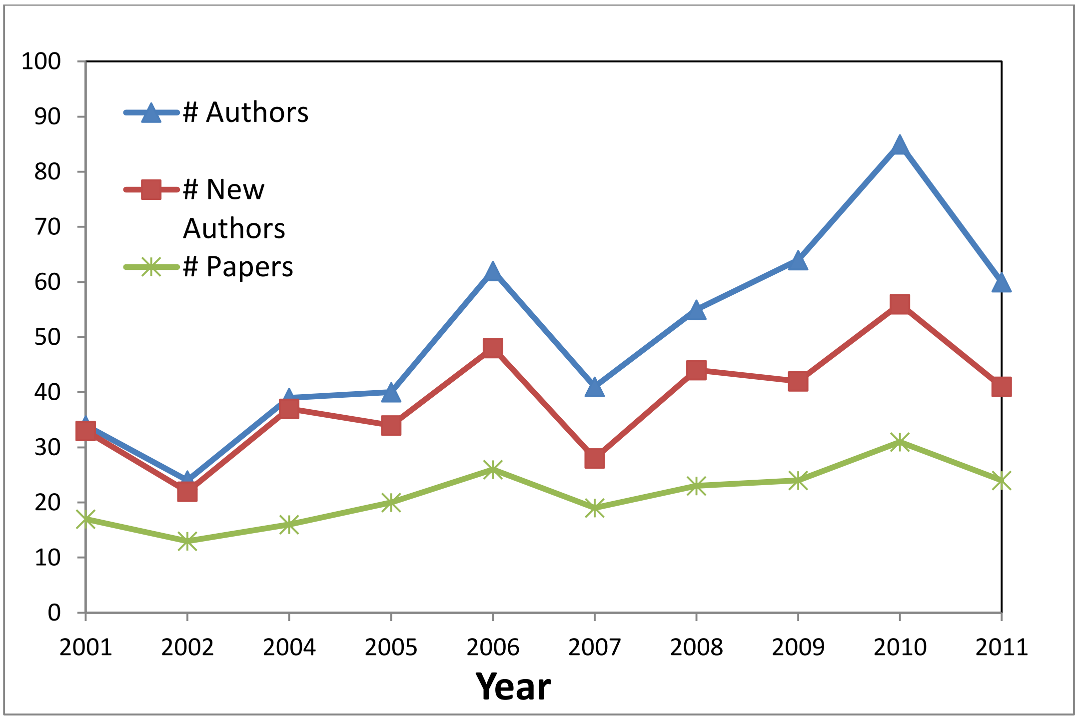

Figure 15: Changes in the Number of Papers, Number of Authors, and Number of new Authors over Time

RQ3.2: How has the structure of the network changed over time?

Over the past decade, the structural characteristics of the CISSE network have evolved in some subtle but distinct ways. During that time new members have joined the network, others have dropped out, while existing members who had not previously collaborated have now began to coauthor papers. Further, new research topics may cause the emergence of new research groups and camps within the network.

In order to analyze the structure of the CISSE network, we divided it into three main temporal boundaries, namely *pre-2005, to 2008* and, *to 2011*. We chose pre-2005 because collaboration between authors was almost non-existent between that time phase. Between 2005 and 2011, we decided to divide this six year time frame into two equal 3-year time period. Two authors are connected temporally, if they published papers together within that time window mentioned above. *Figure 17* and *Figure 18* illustrate the MST structure of the CISSE network for the three time periods. *Figure 16* describes a loosely connected network with just one core group of authors within the network. By 2008, the network structure had swelled to four well-defined groups and about eighteen loosely scattered groups. The network structure continued to develop even further after 2008. In fact, the number of well-defined structures had more than doubled in 2011.

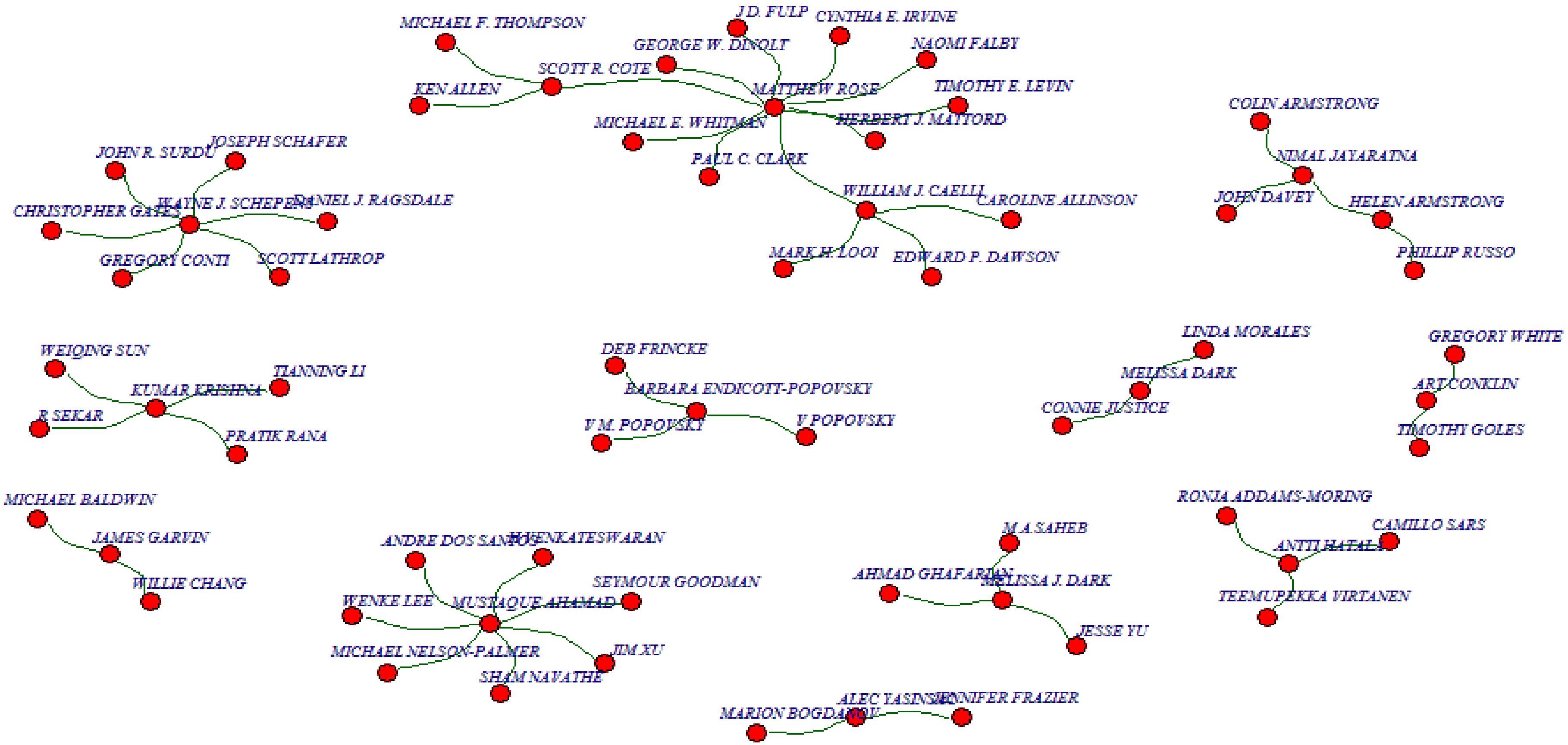

Figure 16: Structure of CISSE Network for pre-2005

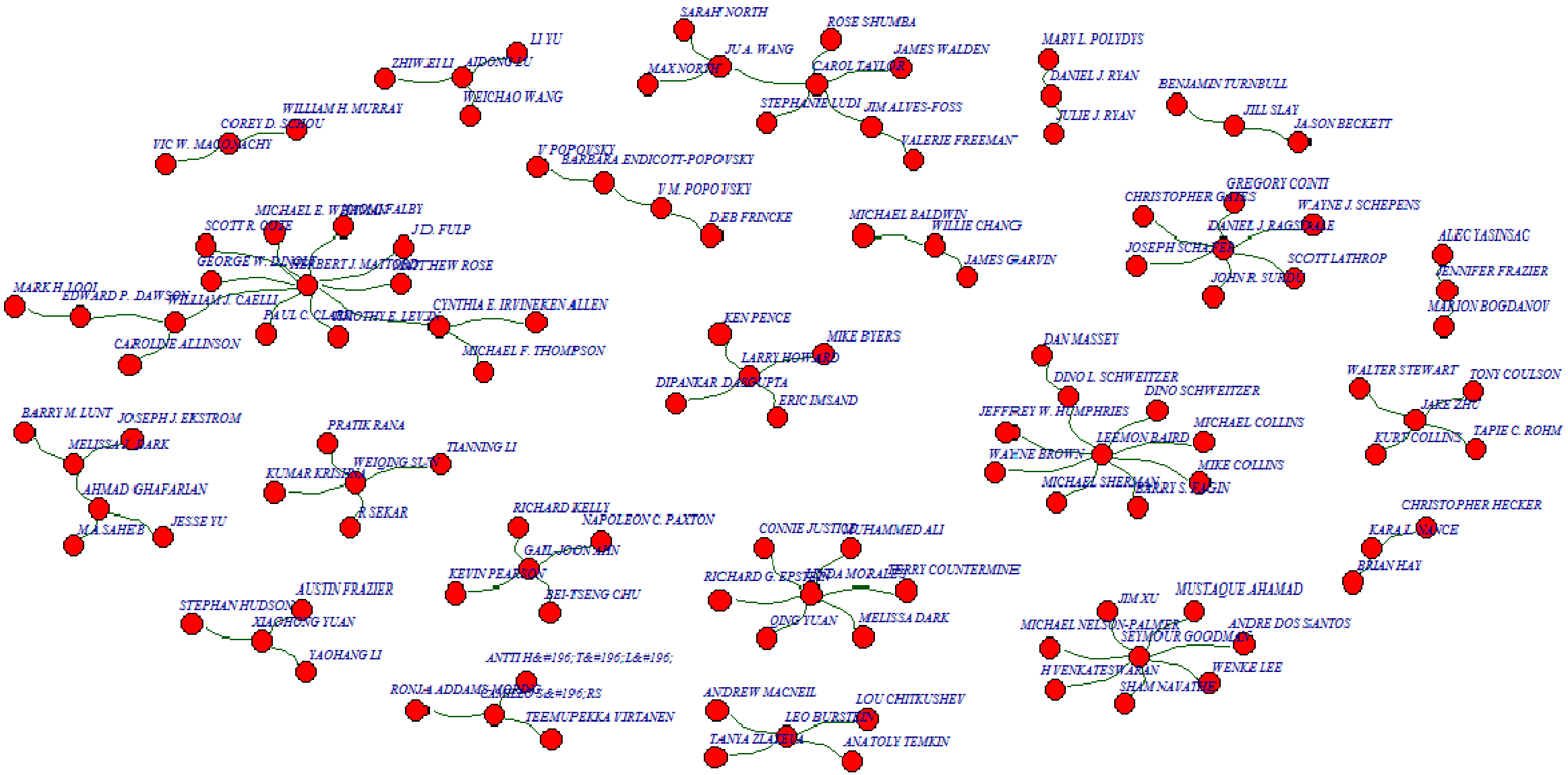

Figure 17: Structure of CISSE Network for 2008

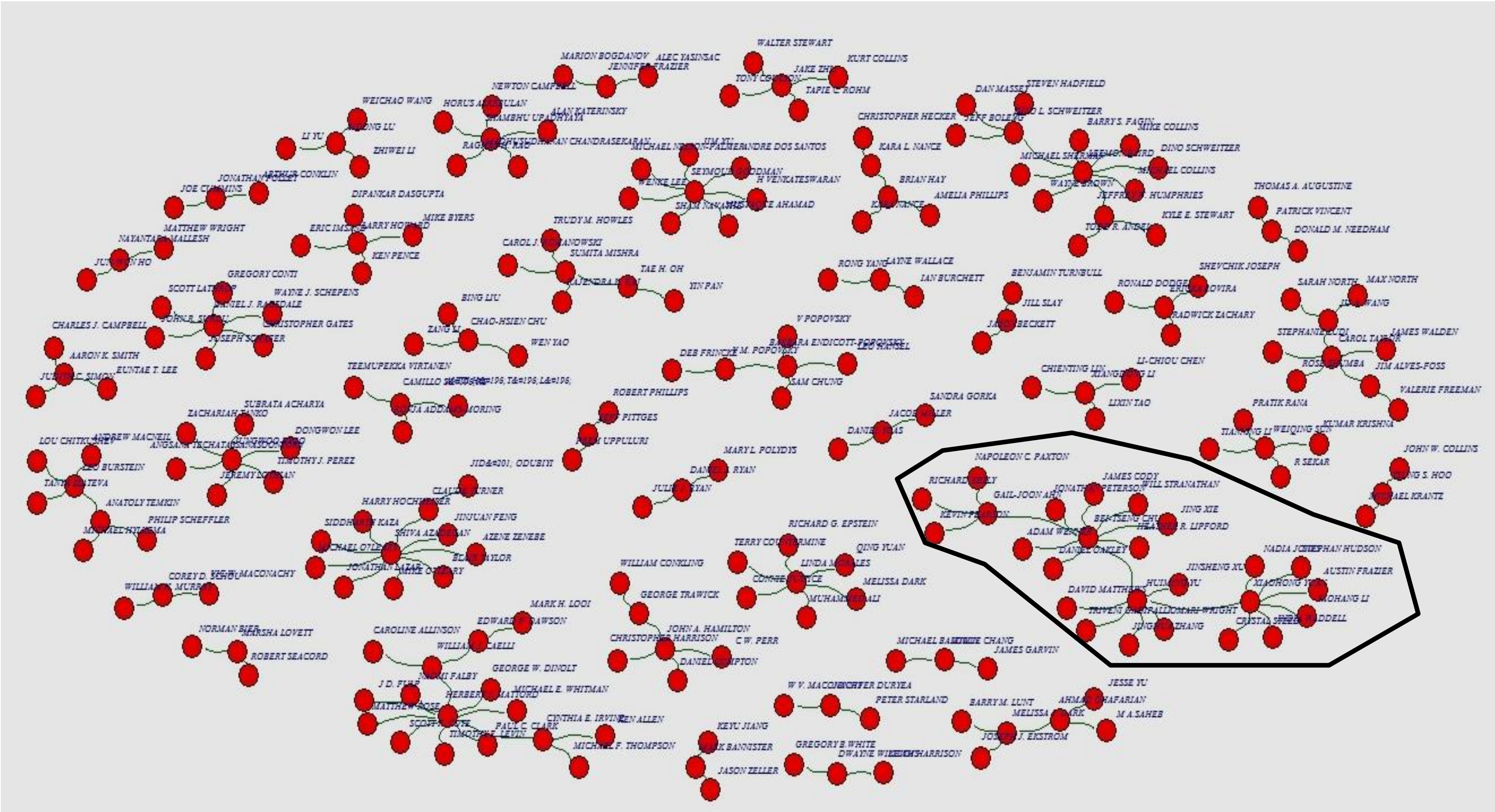

Figure 18: Structure of CISSE Network for 2011

RQ3.3: Are there new generations and stars?

Table 12 and *Table 13* provide a listing of the most productive authors and institutions from 2005 to 2011. Until 2005, collaboration between authors in the CISSE network was almost non-existent primarily because of the recency of the CISSE network (the first published paper was in 2001). In essence, prior to 2005, authors may have developed a wait-and-see attitude to determine the credibility of this new medium as well as perhaps how long this writing avenue will last before committing to publishing papers in it. *Table 12* shows the current star authors in the CISSE network. For a more accurate analysis, as a longitudinal study, it would be interesting to follow these authors for another ten years to observe whether the group dynamics and structure have changed over time. It can also be observed from *Table 12* the star author for the CISSE network hails from the most productive institution in the network.

Table 12: Top Ten Productive CISSE Authors from 2005 - 2011

Author	# of Paper
1. Dino L. Schweitzer	7
2. Richard G. Epstein	7
3. Kara L. Nance	7
4. Brian Hay	6
5. Paul N. Schembari	4
6. Xiaohong Yuan	4
7. Shiva Azadegan	4
8. Blair Taylor	4
9. Jeffrey A. Livermore	4
10. Vic W. Maconachy	4

Table 13: Top Ten Productive Institutions from 2005 - 2011

Institution	# of Paper
1. United States Air Force Academy	26
2. Towson University	23
1. University of North Carolina at Charlotte	19
2. North Carolina A&T State University	17
3. The Pennsylvania State University	16
4. University of Alaska Fairbanks	15
5. The University of Texas at San Antonio	13
6. Auburn University	10
7. Bowie State University	9
8. University of Idaho	9

RQ3.4: How have collaboration patterns changed over time?

For our study, we classify the percentage of papers with multiple authors as the rate of author collaboration. We categorize the percentage of papers that are written by authors from different institutions as the rate of cross institutional collaboration. *Figure 19* below presents the collaboration rates over time for the CISSE network. Clearly, the collaboration rate among authors occurs more frequently and is increasing more steadily and at a steeper slope than cross institutional collaboration. Following remarkable increases between 2005 and 2008, cross institutional collaboration nose-dived in 2009. There was nothing in the literature to explain this phenomenon. However, cross institutional collaboration is showing significant signs of improvement following that twenty-two percent dip between 2008 and 2009.

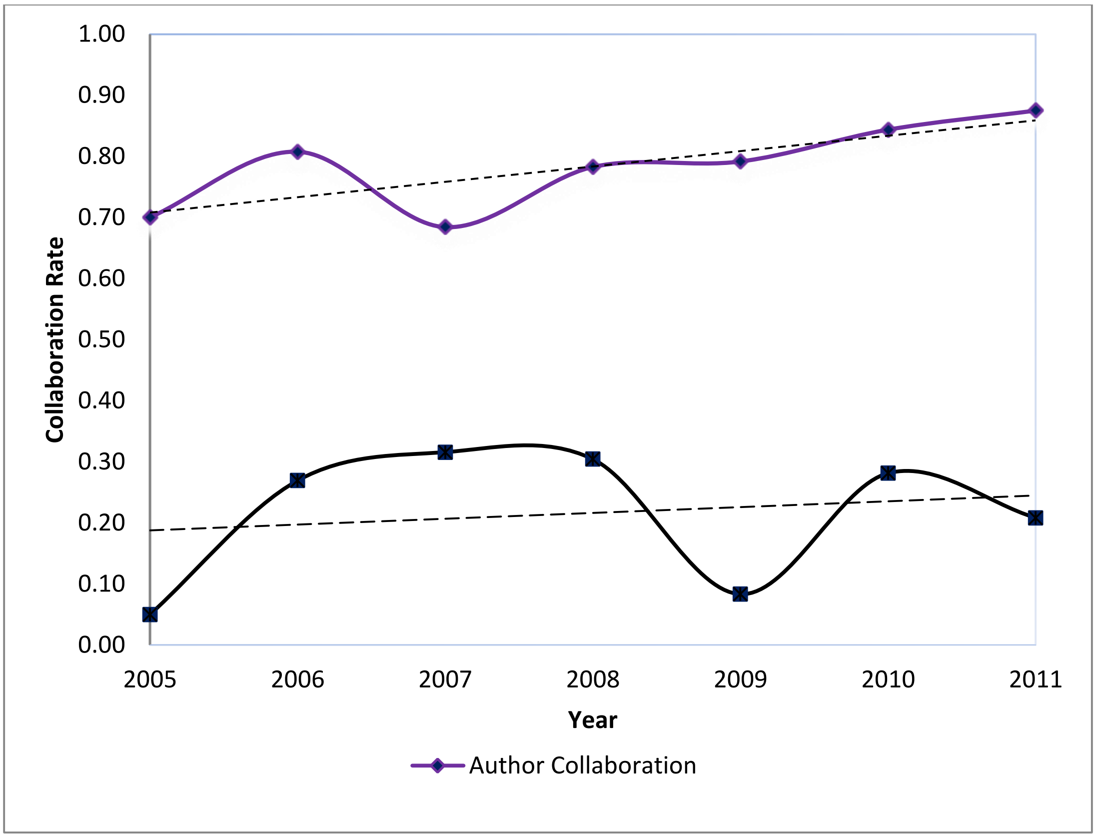

Figure 19: The Changes in Author Collaboration Rate and Institution Collaboration Rate Over Time

4.15.1 Discussion

The results from this study demonstrate a relatively dynamic and overall, an emergent network system of well-connected researchers in the CISSE network community. These moderately cohesive groups of researchers characterize the social identity of the CISSE network. Though the CISSE network has been in existence for a little more than a decade, a leading cluster of prolific researchers provide an essential role in the stability of the community. Since these key researchers have been in the community for a longer period of time, they are more capable of attracting collaborators than would novice researchers. Since the CISSE network continues to develop at a reasonable rate, we anticipate a new generation of rising stars to emerge not just to replace the current key researchers, but to augment the roles played in keeping the cohesiveness of the social community.

4.15.2 Conclusion

Having been in existence for slightly over than a decade, the CISSE network community has evolved into a system of well-connected researchers. Our results underscore the existence of star researchers who either collaborate with other researchers or produce research papers individually. Further, we see a moderate increase in research collaboration at both institutional and cross-institutional levels. Above all, the trend points to an increase in collaboration among researchers, which augurs well for the development, identity and stability of the CISSE community.

Though we observed an increase in the number of individual, institutional collaboration and cross-institutional collaboration of publishers, sadly the number of international publishers has progressively lagged behind. Certainly, to enhance the identity of the CISSE network, the committee must encourage institutional and cross-institutional collaboration at the global level.

A closer examination of the results shows that the most prolific institutions are the government agencies. Coincidentally, the top publishers are also from these institutions. It is imperative that for the continued strength of the CISSE network, a broader cross-section of the community must be incorporated into the system.

PHASE II: THE SOCIAL IDENTITY OF CYBERSECURITY EDUCATION RESEARCH: AN ANALYSIS OF PUBLICATION AND COLLABORATION FROM 1999 – 2013

4.2.1 Introduction

President Barack Obama describes the cybersecurity crisis as "the most serious economic and national security challenge we face as a nation" [126]. Over the past few decades, the national infrastructure, including military, financial, power, and telecommunication systems, has become increasingly reliant on computer software and networks. As a result, cybersecurity education and research has become a major area of concern for the government, academia and industry. Academia has focused on preparing college students and information technology professionals in computer security and recently a great deal of attention has been paid to curriculum development guidelines like the ACM CS 2013[127] and government agency designations like the National Security Agency/Department of Homeland Security Center of Excellence in Cyber Defense and Cyber Operations[128] focusing on college curricula with an emphasis on program articulations [105], integration of information assurance in the computer science curriculum [129], and the effectiveness of these classes as well as the degree to which these students are trained to handle security and privacy challenges [105]–[108].

A number of scholars and practitioners from top institutions from around the world, including public and private sectors have written extensively on Cybersecurity education. Some of them have presented models of pedagogical research to enhance and expand the needs of the discipline [130], [131]. Other researchers have designed fundamentally sound techniques for fostering IAE and SE based on established scientific principles [132]–[134]. Further, practitioners have collaborated within and across institutions and geographical boundaries. The collaborations among these researchers represent a complex network structure with complete distinctive structural characteristics. Social Network Analysis (SNA) methods can be used to study the levels of collaboration in the community, identify key institutions, individuals, and other stakeholders and study the growth of the cybersecurity education discipline.

The fundamental point of this study is to analyze the research collaborations of Cybersecurity researchers who presented papers in IEEE Xplore and the ACM digital libraries between 1999 and 2013. Given that IEEE Xplore and ACM provide a powerful resource for IAE and security education publication, we believe those areas provide an idyllic test bed to evaluate the community using SNA techniques.

We explore the following research questions: (1) what are the patterns of interactions between members of the IAE community? (2) are there core sets of active members who sustain the network?, and (3) how has the network evolved over time? The next section contains a review of social network analysis literature. Section 3 presents our research design and dataset and methods. Section 4 presents and discusses the results and Section 5 concludes and presents future directions.

Section 4.2.2: Literature Review

See Section 4 of Literature Review Chapter

Section 4.2.3: Research Design and Data Set

The overall existence of the IEEE Xplore and the ACM digital libraries are to provide a powerful resource for discovery and access to scientific and technical content published by the IEEE and its publishing partners (IEEE website) and to deliver resources that advances computing as a science and a profession; to enable professional development; and to promote policies and research that benefit society, respectively (ACM website). The IEEE Xplore digital library which is considered to be the world's largest professional association dedicated to advancing technology, is available to the public and to organizations for free as well as for a subscription basis. The library covers topics from all facets of Computer Science and reflects national and international participation of leading figures in government, industry, and academia.

The general information available to the public includes the paper, the title of the article, the author(s) of the article, the author(s) organization affiliation, keywords, metrics, date of publication, sponsor as well as the article abstract. Similar to IEEE, the ACM Digital Library also covers a variety of topics in computer science and is available to the public as well. The information available to the public is considerably the same as what's available in the IEEE library.

Using a crawler we developed in R, we scraped the web pages from both of the above mentioned websites using the following criteria in the queries: "Cybersecurity education ", "Security Education" and "Cyber Security Education". Then we downloaded the majority of the general information available to the public including the article abstract, author(s) information, article title, publication date and proceeding for all papers that were published between 1999 and 2013.

All the data for the period were downloaded and saved to a database where the information analysis was prepared. Altogether, we downloaded data for 12, 083 hits for a total of 12, 113 unique authors from over 3, 661 institutions throughout the world with 3, 405 article collaborations from the IEEE Xplore and ACM Digital Libraries *(see Table 14 below)*. The analyses for this study were done using six well popular Social Networking Analysis (SNA) tools: namely UCINET, PAJEK, R, igraph, NetDraw, and Gephi.

Table 14: Summary of Hits

Summary Area	**Total #**
Record / Hits	12, 083
Unique Authors	12, 113
Unique Institutions	3, 661
Author Collaborations	3, 405
Unique Institutional Collaborations	43
Unique Titles	4, 254

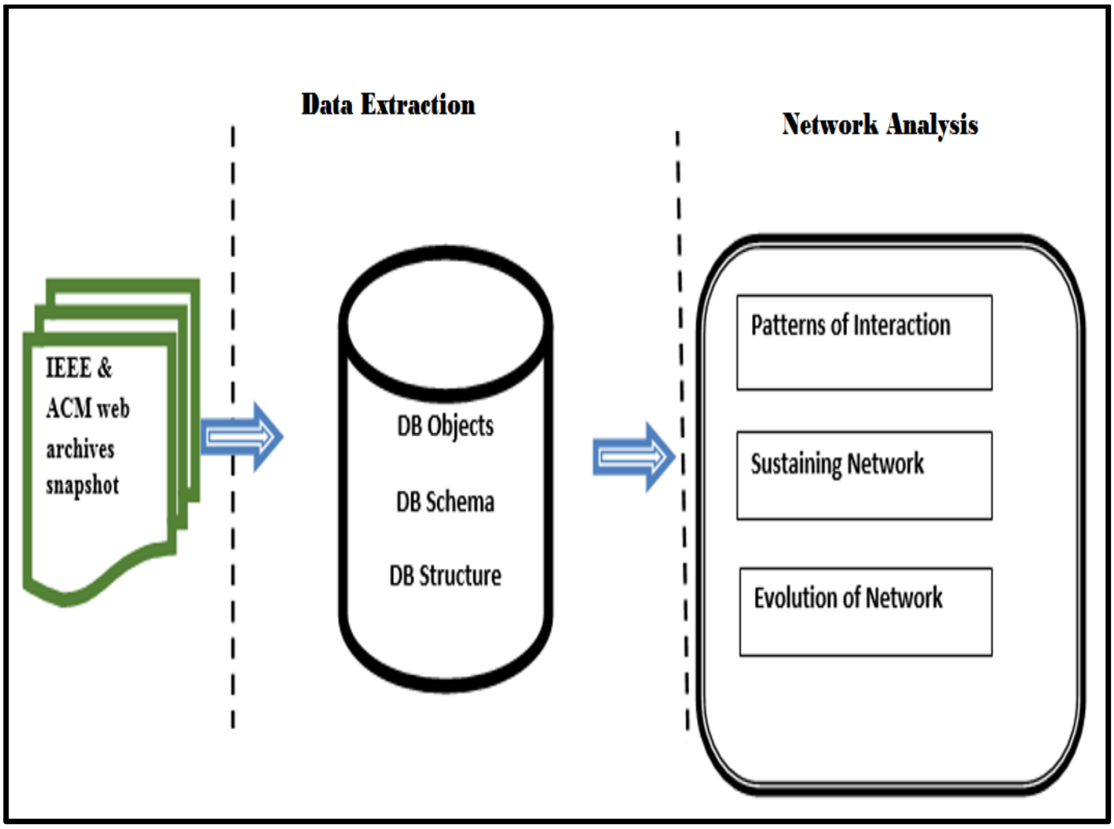

Figure 20: IEEE - ACM Research Design

Our focus is on analyzing research collaboration of Cybersecurity education researchers who coauthored IEEE and ACM papers. Specifically, we examine the structure of their associated networks to determine the social identity of the discipline using social network analysis (SNA). The following are our primary research questions:

R1: What are the patterns of interactions between members of the IAE community?

When members of the research community collaborate on research projects, a system of networks is inevitably formed. The more frequent the interactions among members, the stronger the ties between the community which ultimately enhances information diffusion within the IAE community. There are three sub-questions:

R1.1: Are IAE researchers connected?

R1.2: What structural characteristics does the collaboration network have?

R1.3: Are there subgroups in this community?

R2: Are there core sets of the "critical mass of active members who sustain the network"?

Here we try to determine the research leaders, who co-authored with those leaders as well as their institution. This research question is sub-divided into four sub-sections:

R2.1: Who are the most eminent researchers in the community?

R2.2: Are these star researchers critical to holding the community together?

R2.3: Which institutions are the most productive?

R2.4: What are the cross-institutional collaboration patterns?

R3: How has the network evolved?

We try to examine what changes have taken place to the research community over time. Specifically, has the discipline been able to attract new members on a consistent basis since its inception. We analyze the evolution of the community in the following manner:

R3.1: Has the community been able to absorb new members?

R3.2: How has the structure of the network changed over time?

R3.3: Are there new generations and stars?

R3.4: How have collaboration patterns changed over time?

Section 4.2.4. Results and discussion

4.2.4.1 Structural characteristics

Our results revealed the IEEE and ACM network is weakly connected. The IEEE and ACM network collaboration's largest component accounting for 21.53 percent of all the authors who have published articles in IEEE and ACM network. Seventy four percent of the authors collaborated on projects within groups comprising of 2 or more members. Only 6.8 percent of the IEEE and ACM authors have never collaborated with anyone except themselves. There were multiple collaborations among the same authors. In fact, collaborations between two authors occurred one hundred and sixty two times with the highest number of collaborations between any pair being five. The average number of researchers working on an article is three. See Figure 2 below.

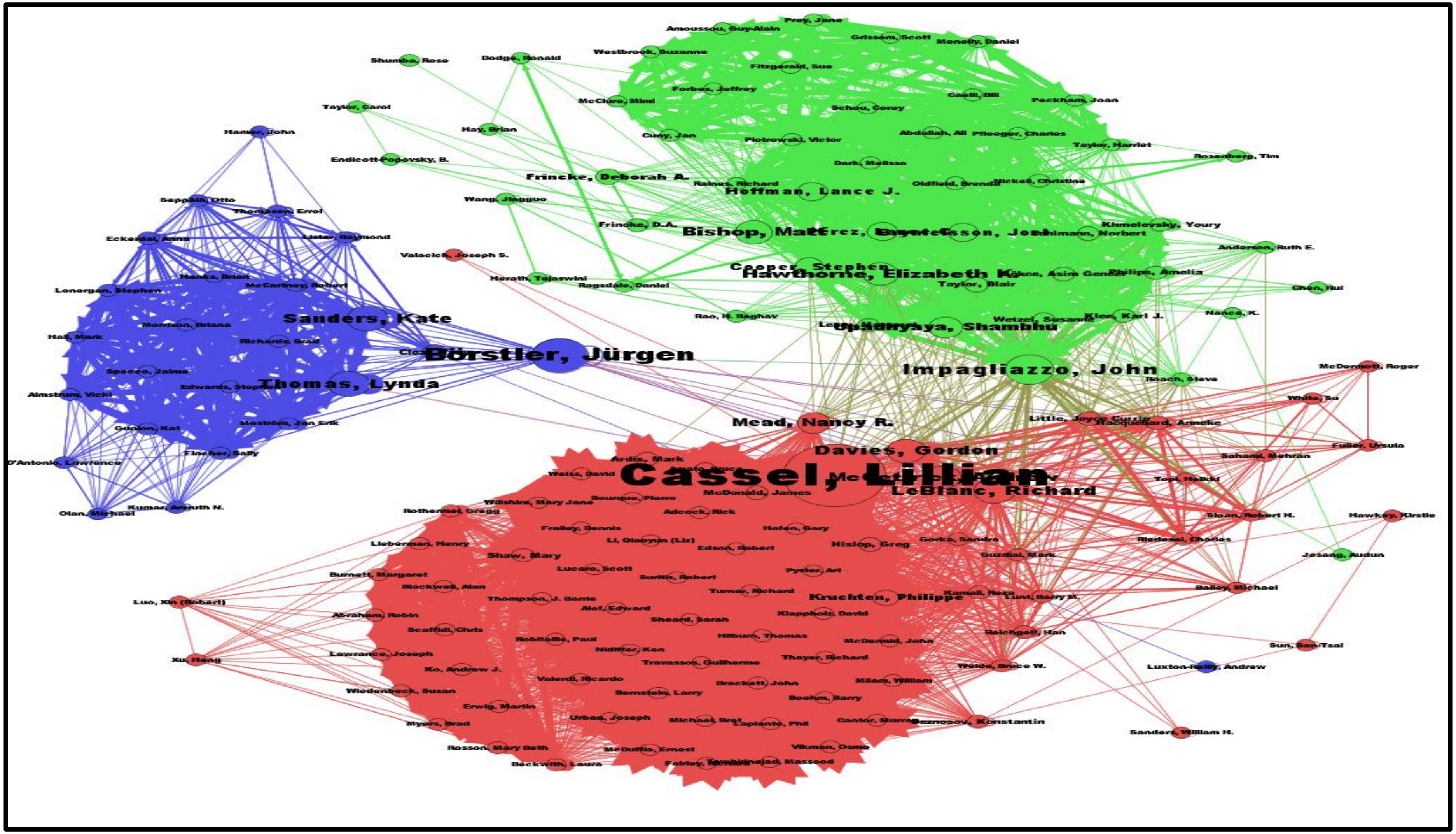

Figure 21: Network Components of Authors

RQ1.1: Are IAE researchers connected

Two of the measures, *distance* and *efficiency*, can be used to determine information diffusion through the network. Whereas distance is defined by the shortest path between two nodes [135], the efficiency of a network is the sum of the inverses of lengths of all-pairs shortest paths [124]. We found that the average distance in the giant component was 2.6 while the closeness or efficiency was 0.402. In order to compare the random network with its counterpart, we generated thirty random networks of the same size and average degree with those of the giant component. We found a resulting mean distance of 1.8 and an efficiency of 0.566. Essentially, the researchers in the IEEE and ACM community and the random network are not identical. The average distance of the IEEE and ACM community is higher than the random network while the efficiency is lower than the random network.

RQ1.2: What structural characteristics does the collaboration network have?

Networks structures vary depending on their type of collaboration. For instance, in a random collaboration, researchers select collaborators on a random basis. On the other hand, with a scale-free structure, a few active members disproportionately dominate the activity within the network. We anticipate the IEEE and ACM community to be of a scale-free network structure.

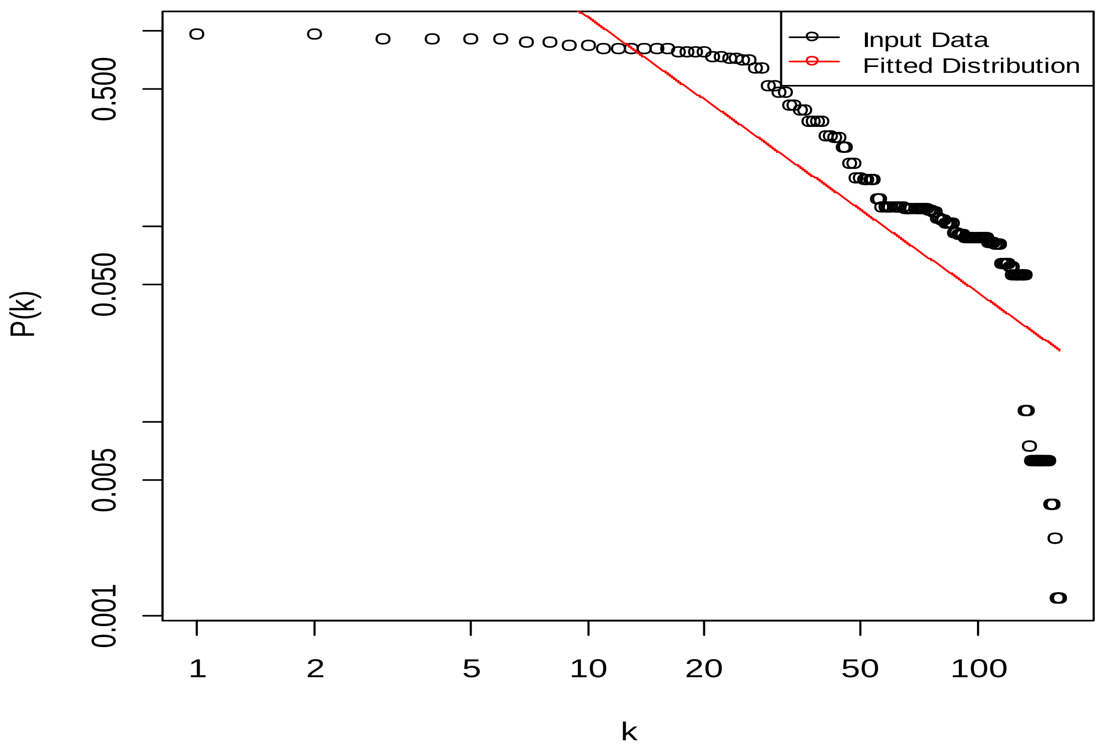

Figure 22: Cumulative Degree Distribution of IEEE and ACM Network

Figure 22 illustrates the cumulative degree distribution of the IEEE and ACM network. As can be observed, the diagram clearly represents a distinctive scale-free characteristic with a few authors having a large number of collaborations and many authors with a significantly small number of collaborations. The diagram depicts that the IEEE and ACM network is not a true power-law relationship because the tail shows a steep decline in the number of collaborators. The findings indicate that researchers have been able to attract other researchers from other institutions and countries on a more readily basis. Further, the research showed that cross institution collaboration occur much more spontaneously among large public institutions in the United States of America than the smaller and private counterparts.

RQ1.3: Are there subgroups in this community?

Since the researchers in the IEEE and ACM community typically share a common interest, exclusively security, we can use cluster analysis to determine the extent those researchers form subgroups in the network. To find whether the network clusters are true subgroups, we analyze the collaboration using a network analysis statistical tool called cluster coefficient. With random networks it is highly unlikely to obtain subgroups within the network because each node has the same probability to connect to other nodes. Therefore, the cluster coefficient for random network is characteristically rather small. The cluster coefficient for the IEEE and ACM network was 10.207 compared to the 0.235 of its random subgroup. Essentially, this means strong cliques are form among researchers in IEEE and ACM network. This could be attributable to fact that most of the authors in IEEE and ACM network collaborated on projects within groups of 2 or more members. Our findings indicate that there are evidence of a "small word" in the IEEE and ACM community because of the overall short distance for the subgroups and a significant higher cluster coefficient than random subgroup.

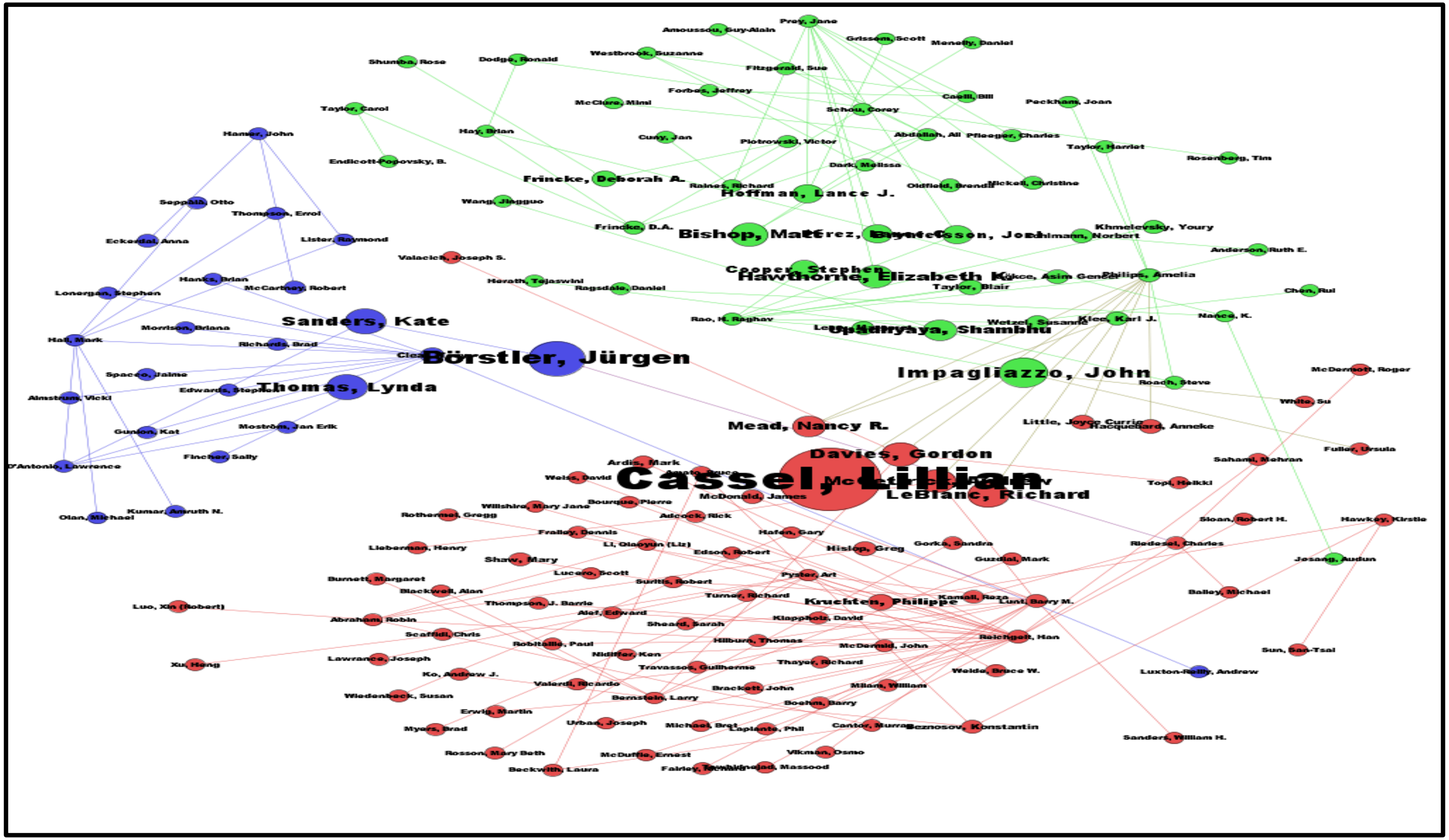

Figure 23: Minimum Spanning Tree of IEEE and ACM Network

RQ2: Are there core sets of the "critical mass of active members who sustain the network"

RQ2.1: Who are the most eminent researchers in the community?

Eminent researchers are those researchers who may have published several papers either on their own or perhaps with many collaborators. We have decided to analyze an author's productivity based on three factors used by Xu and Chau, namely; *normal*, *straight* and *adjusted ranks*. Normal rank is determined based on the number of papers an author publishes. On the other hand, straight rank counts only the papers of which the author is the first author. Adjusted ranks is built on collaborative strength of which the author receives 1 for a paper with only two authors but a less than 1 value with papers with more than two authors [136]. *Table 15* below lists the top authors in each productivity category. The combination of all three productivity ranking should give an unambiguous representation of who the most productive researchers are in the IEEE and ACM community. The results of the study show that some key persons exist in the network because they rank high in each category. For instance, Matt Bishop, Elisa Bertino and Ninghui Li rank high in each category. The results for the study denote a large, sparsely distributed and isolated network with a power-law of 0.005 and with an average clustering coefficient of 0.001 and R-squared of 0.00034 or 0.03 percent. This suggests that there is practically no correlation between authors and the numbers of papers published. Hence, our findings prove that just a very few authors have published a large number of papers.

Table 15: Author Productivity Rank

Name	# Of Papers	Name	# 1st Authored Papers	Name	Adjusted Score
1. Mikhail J. Atallah	39	1. Matt Bishop	6	1. Matt Bishop	12.82
2. Matt Bishop	28	1. Wasim Al-Hamdani	6	2. Ninghui Li	6.74
3. Elisa Bertino	26	1. Elizabeth Hawthorne	6	3. Elisa Bertino	6.67
4. Mario Piattini	21	4. Ed Crowley	5	4. Eugene H. Spafford	6.2
5. Raghav Rao	17	4. Mich Kabay	5	5. Melissa Dark	5.65
5. Ninghui Li	17	6.Peter Neumann	4	6. Mikhail J. Atallah	5.58
7. Eugene Spafford	15	6. John Gorgone	4	7. Brian Hay	4.28
8. Xiang-Yang Wang	14	8. Frank Katz	3	8. Xiang-Yang Wang	4
8. Elizabeth Hawthorne	14	9. Eugene H. Spafford	2	9. Cynthia E. Irvine	3.42
10. Brian Hay	13	9. Cynthia Irvine	2	10. Doug Jacobson	3.33

Table 16: Author Centrality

Degree	# of Collaborators	Closeness	Betweenness
1. Mikhail J. Atallah, Purdue University, USA	38	1. Mikhail J. Atallah, Purdue University, USA	1. Han Reichgelt, University of South Florida, USA
2. Elisa Bertino, Purdue University, USA	26	2. Matt Bishop, University of California - Davis, USA	2. Raghav Rao, State University of New York at Buffalo, USA & Sogang University, Korea
3. Matt Bishop, University of California - Davis, USA	21	3. Elisa Bertino, Purdue University, USA	3. Mercan Topkara, Purdue University, USA
3. Mario Piattini, University of Castilla-La Mancha Camino de Moledores, Spain	21	4. Raghav Rao, State University of New York at Buffalo, USA & Sogang University, Korea	4. Andrew Mc Gettrick, University of Strathclyde, UK
5. Ninghui Li, Purdue University, USA	17	4. Elizabeth Hawthorne, Union County College, USA	5. Mohamed Shehab, Purdue University, USA
5. Raghav Rao, State University of New York at Buffalo, USA & Sogang University, Korea	17	4. Eugene H. Spafford, Purdue University, USA	5. Nancy R. Mead, Carnegie Mellon University, USA
7. Xiang-Yang Wang, Tsinghua University, China	13	4. Han Reichgelt, University of South Florida, USA	5. Ursula Fuller, University of Kent, UK
7. Eugene H. Spafford, Purdue University, USA	13	4. Raghav Rao, State University of New York at Buffalo, USA & Sogang University, Korea	5. Kirstie Hawkey, University of British Columbia, Canada
9. Brian Hay, University of Alaska, USA	10	10. Daniel Ragsdale, United States Military Academy, USA	5. Radu Sion, Stony Brook University, USA
10. Elizabeth Hawthorne, Union County College, USA	8	10. Barbara Endicott-Popovsky, University of Washington, USA	10. Robert Crossler, Mississippi State University, USA

To determine the prominence of a researcher, we have decided to use degree centrality of the IEEE and ACM network. The degree of a researcher is the number of collaborators he or she has. In addition, we utilized two other commonly used centrality measures: *closeness* and *betweenness*. Both of these measure the reachability of a person within a network. Betweenness, measures the extent to which one node lies between other nodes in the network. For instance, how significant a person is to the transmission of information through a network as well as the extent to which a person may control the flow of information due to his or her position in the communication network can be traced back to his or her betweenness. Closeness, on the other hand, measures the distance from one node to all other vertices within the network. If you want to measure indirect contact to neighbors in the entire network, then closeness is the preferred measure. The significance of the vertex in the network analysis is best captured using betweenness centrality. *Table 16* above shows mostly a low level of closeness among the researchers, except for the top ten authors. The majority of the ten top productive authors have the most power and influence in this network. As can be observed, none of the top performing authors is associated with the betweenness centrality. The results suggest that no single person in the network can exert their control on any other individual in the network although the top producing authors have the highest influence. This would seem to indicate perhaps isolated pairs of vertices in the network which does not contain a giant component.

RQ2.2: Are these star researchers critical to holding the community together?

Unlike a structural cohesive environment in which everyone plays an equally important role in the network community, actors in a scale-free network have clearly defined roles which enable network connectivity. To determine the central roles of key researchers in the network, we decided to perform a *"network robustness test"* [65] to observe how the network would alter if those main researchers are missing. Fundamentally, our hope is to identify an optimal set of key nodes. If we progressively delete those key nodes from the network, at some point the network should completely cripple the network. Further, since well-connected nodes are likely to possess a great deal of information, and who, by virtue of their connections are in a position to influence others, breaking the connections would ultimately cause members to have a difficult time to communicate. Moreover, removing key nodes would inevitably cause the network to become small clusters leading to an extremely small average node distance.

We first identified the key researchers in the network using KeyPlayer1[125] and presented the diagrams using NetDraw. The top ten key players in the IEEE and ACM networks are presented in *Figure 17* below. We then removed the nodes for the key players in the network not only to increase the distance between some pairs of nodes but to completely disconnect them as well. *Figure 24* shows the disconnectedness and fragmentation of the IEEE and ACM once the key players were removed. The results of the removal of those key players left us with a non-cohesion measure of 0.962 percent. It should be noted here that the closer the non-cohesion measure is to 1 the greater the optimization and the smaller the number of clusters within that network.

Table 17: Key Players

Author
Matt Bishop
Kara Nance
Daniel Ragsdale

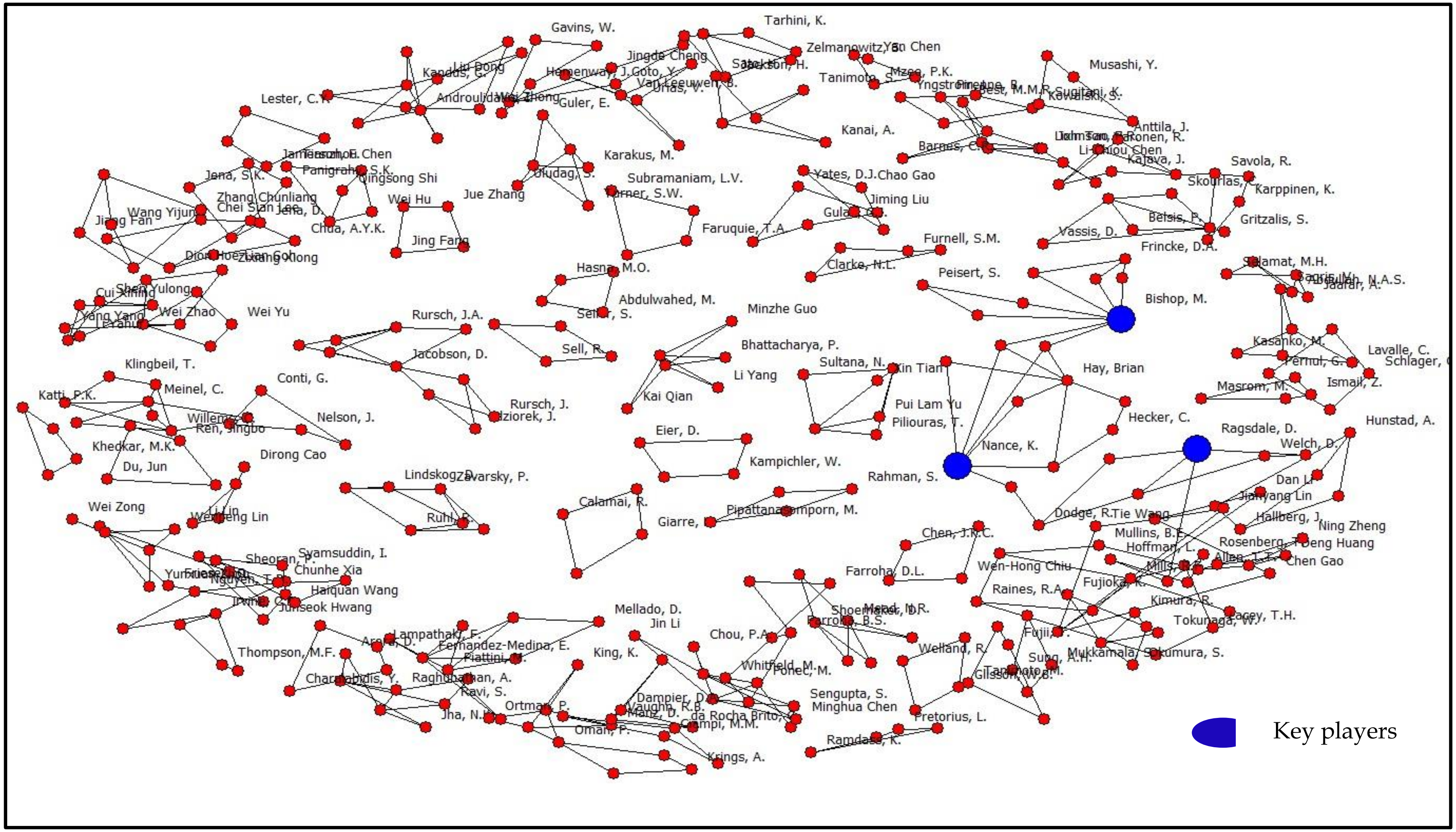

Figure 24: Key Players in IEEE and ACM Network

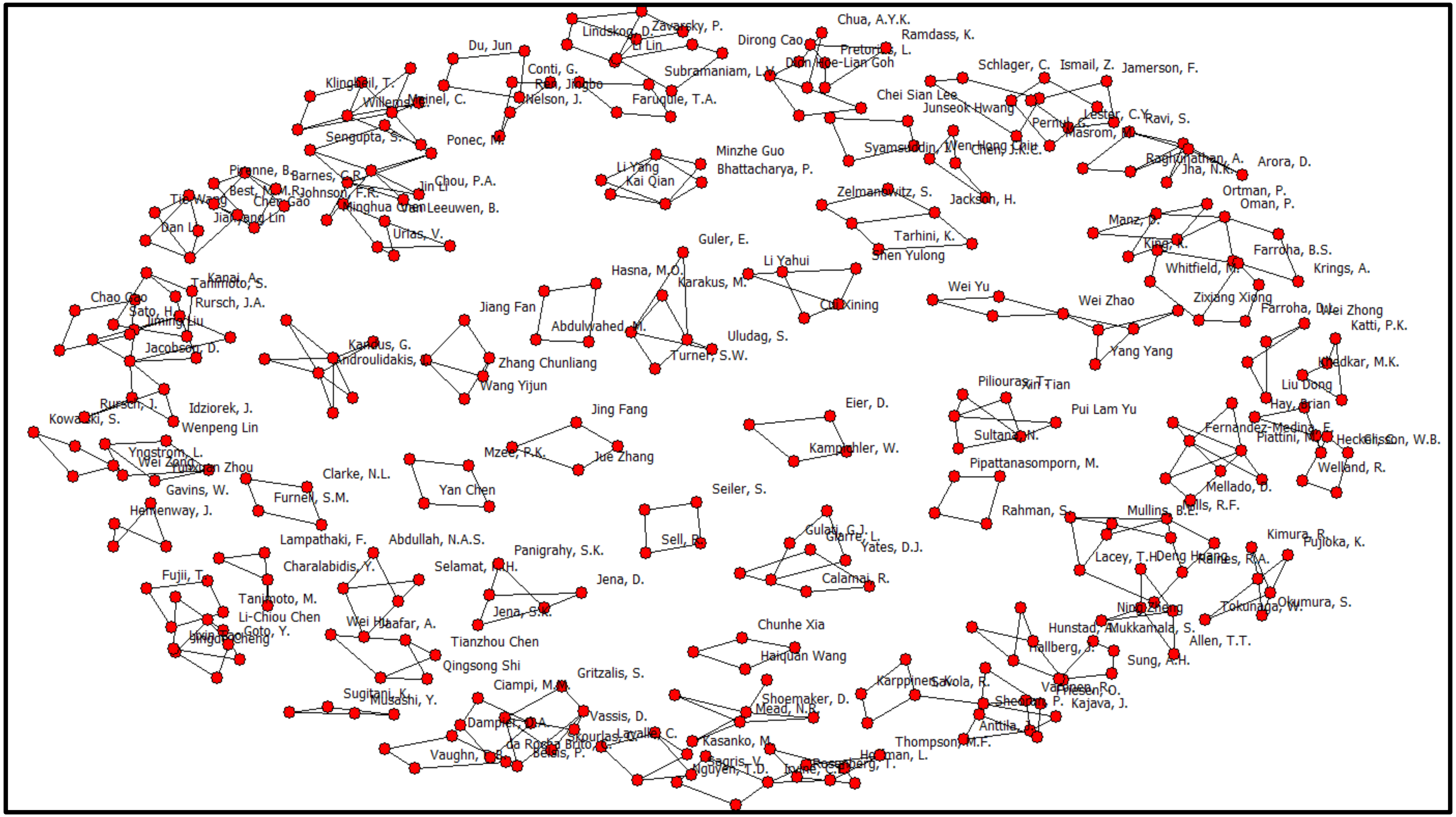

Figure 25: IEEE and ACM Network after Deleted Key Players

RQ2.3: Which institutions are the most productive?

Table 18: Top 10 Productive Institutions

Institution	**# of Papers**	**Institution**	**# of First-Authored Papers**
1.University of California, CA, USA	70	1.Iowa State Univ., Ames, IA, USA	9
2. Carnegie Mellon University, PA, USA	30	2. Sandia National Labs., Albuquerque, NM, USA	7
3. US Military Academy, West Point, NY, USA	26	3. North Carolina A&T State University at Greensboro	6
4. Iowa State Univ., Ames, IA, USA	23	4. Hasso Plattner Institute, Univ. Potsdam, Germany	5
5. Georgia Institute of Technology, GA, USA	20	4. Instituto Superior de Engenharia de Coimbra (Coimbria Institute of Engineering), Coimbria, Portugal	3
6. Naval Postgraduate School, CA, USA	19	4. Jožef Stefan International Postgraduate School, Ljubljana, Slovenia	3
7. Sandia National Labs., Albuquerque, NM, USA	12	4. Booz Allen Hamilton, Inc., Los Angeles, CA, USA	3
7. University of Southern California, CA, USA	12	5. QinetiQ, Farnborough, UK	2
9. California State Polytechnic University, CA, USA	11	5. California State Polytechnic University, Pomona, CA	2
10. North Carolina State University, North Carolina, USA	9	5. United States Military Academy	2

We measured the universities' productivity using both normal and straight rank. *Table 18* above presents the top ten productive institutions. The Iowa State University in Ames, the North Carolina A&T State University in Greensboro and Hasso Plattner Institute University in Potsdam, Germany are high in both categories. The United States Military Academy in West Point, New York and the California State Polytechnic University in Pomona, California are also high in number of papers published. The results from the study illustrate that all except three of the top ten institutions from either group is from another country other than the United States of America.

RQ2.4: What are the cross-institutional collaboration patterns?

The overall existence of the IEEE *Xplore* and the ACM digital libraries are to provide a powerful resource for discovery and access to scientific and technical content published by the IEEE and its publishing partners (IEEE website) and to deliver resources that advances computing as a science and a profession; to enable professional development; and to promote policies and research that benefit society, respectively (ACM website). Since there are many facets in the publishing arena, researchers, who ultimately have that common interest, should therefore find it relatively straightforward to collaborate not only within their institutions but also on a regional and international scale as well.

An analysis of the researchers who collaborated with other researchers that were not a faculty member of their institutions was examined. Using that data, a network diagram (*Figure 26);* a giant component (MST) (*Figure 27)* and a minimum spanning tree (MST) (*Figure 28*) were constructed on all researchers who participated in cross-institutional IAE papers. An institution ***A*** is considered to be connected to institution ***B*** if authors from both institutions have collaborated in at least one paper. The network collaboration links are based on the existence of an alliance in publishing rather than the frequency of their publication.

The giant component for the collaboration network comprise of 120 nodes which accounts for 3 percent of the institutions that have published in IEEE Xplore security and ACM networks. There are 4883 links among these institutions. On average, an institution collaborates with 2.52 other institutions. The top two collaborative institutions are the University of California in the United States of America and United States Military Academy, West Point, New York, which have collaborated with 16 and 10 other institutions, respectively. Coincidentally, 80 percent of the collaborative institutions are within the United States of America.

In *Figure 26*, *Figure 27* and *Figure 28* below, the collaborative network and MST for the institutions are presented. The ten most prolific institutions are labeled with their names. It can be observed from *Figure 26* that there is a loose concentration of institutions that collaborate in the IEEE and ACM network. The majority of the collaborating institutions are located in the United States. Though some cross collaboration exists between institutions across country borders, the collaboration primarily occur between institutions within the same regions.

The minimum spanning tree (MST) for the cross-institutional collaborative network for IEEE and ACM researchers is rather interesting. It shows that collaboration exists between institutions on a highly global basis. For instance, the faculty of Curtin University in Australia collaborated with their counterparts at the

University of California, while the faculty at the University of Maryland collaborated with other faculty from Beijing Technology & Business University in Beijing, China. Further, cooperation between institutions within the United States is not limited to geographic proximity. For example, the University of Maryland on the east coast of the United States teamed up with the University of California on the west coast. Finally, a close examination of the MST will indicate that not only are the main collaborative institutions located in the United States of America, the majority of these institutions are large public universities in the United States of America.

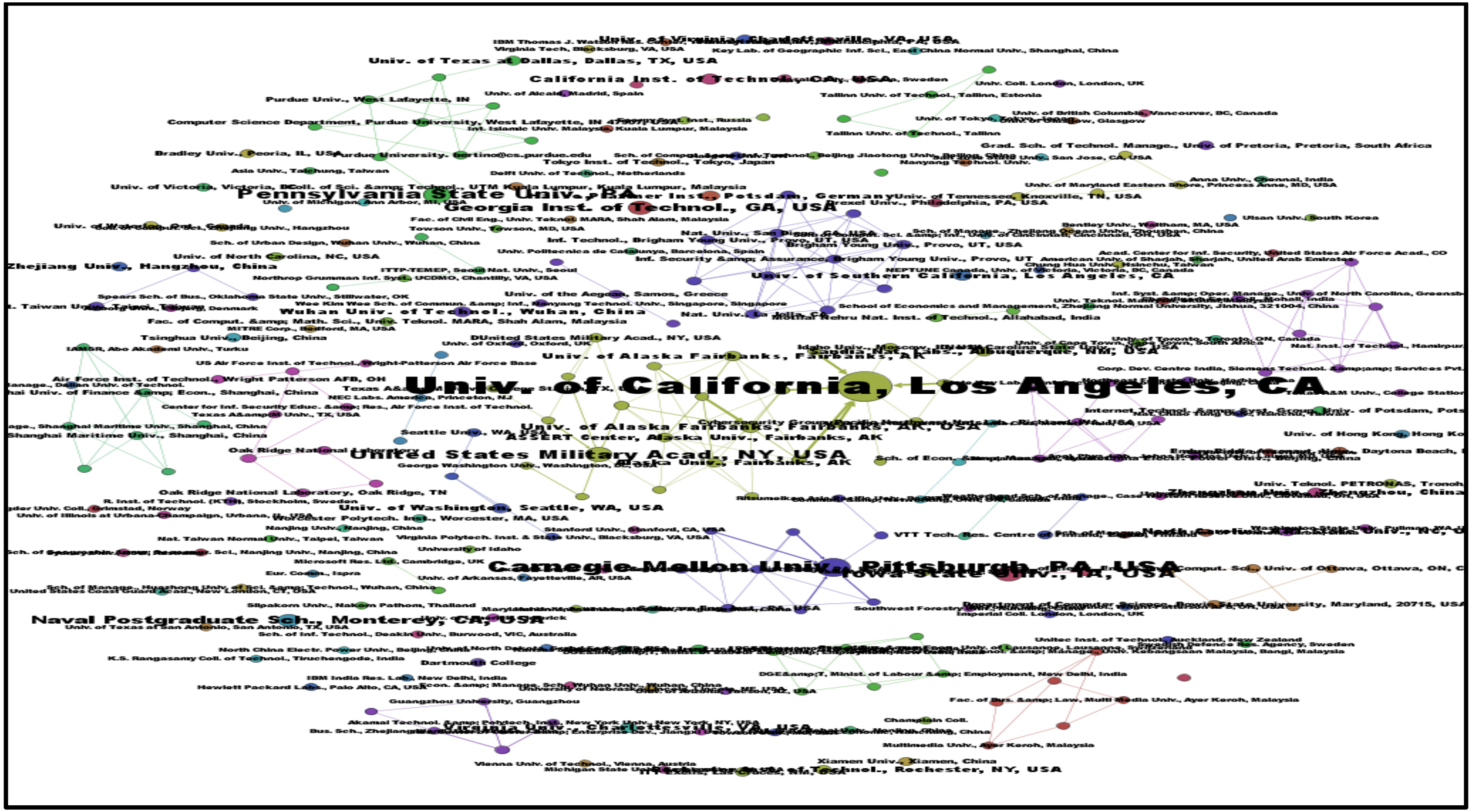

Figure 26: Cross-institutional Collaboration Network

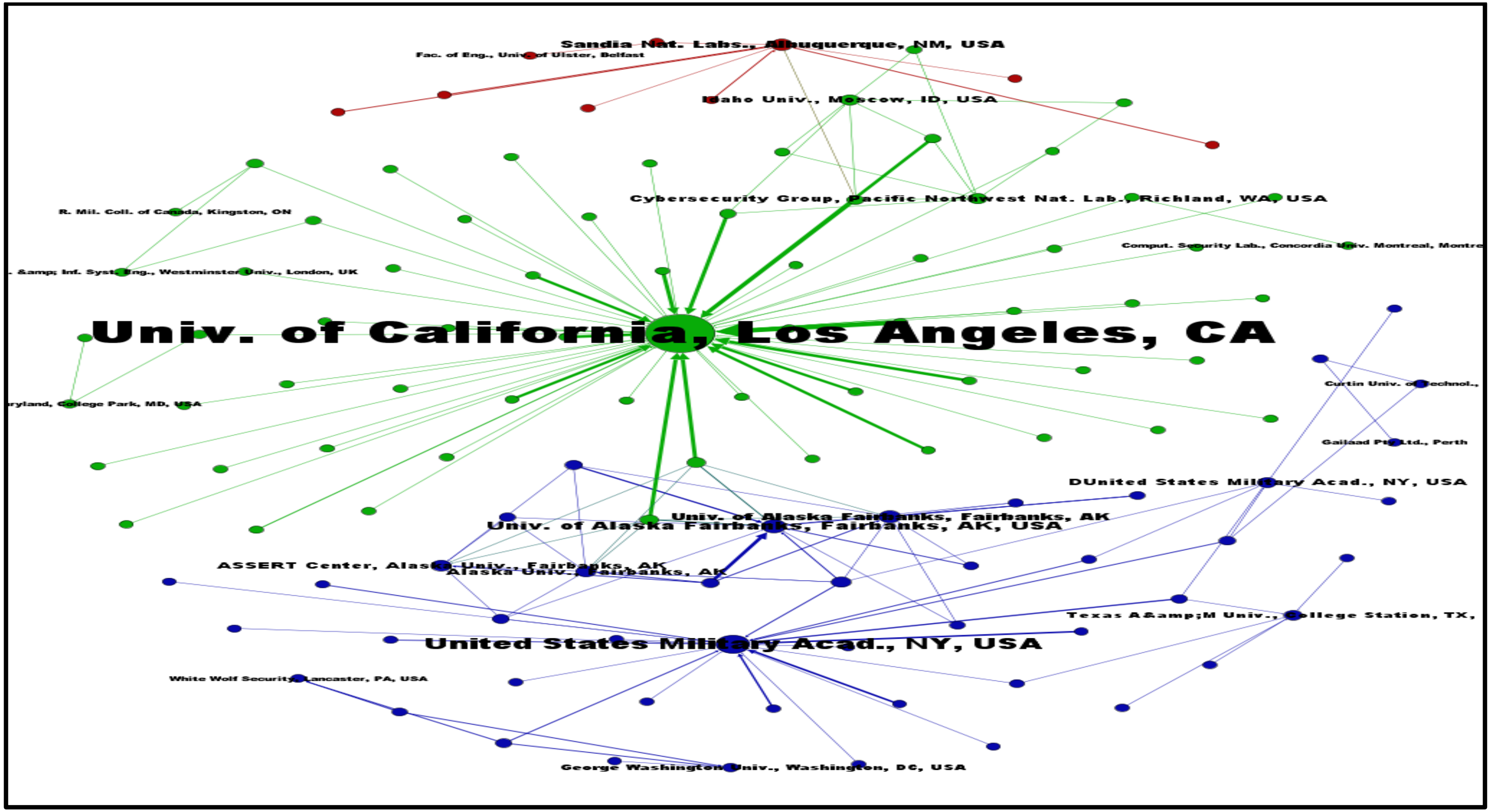

Figure 27: Cross-institutional Collaboration Network (Giant Component)

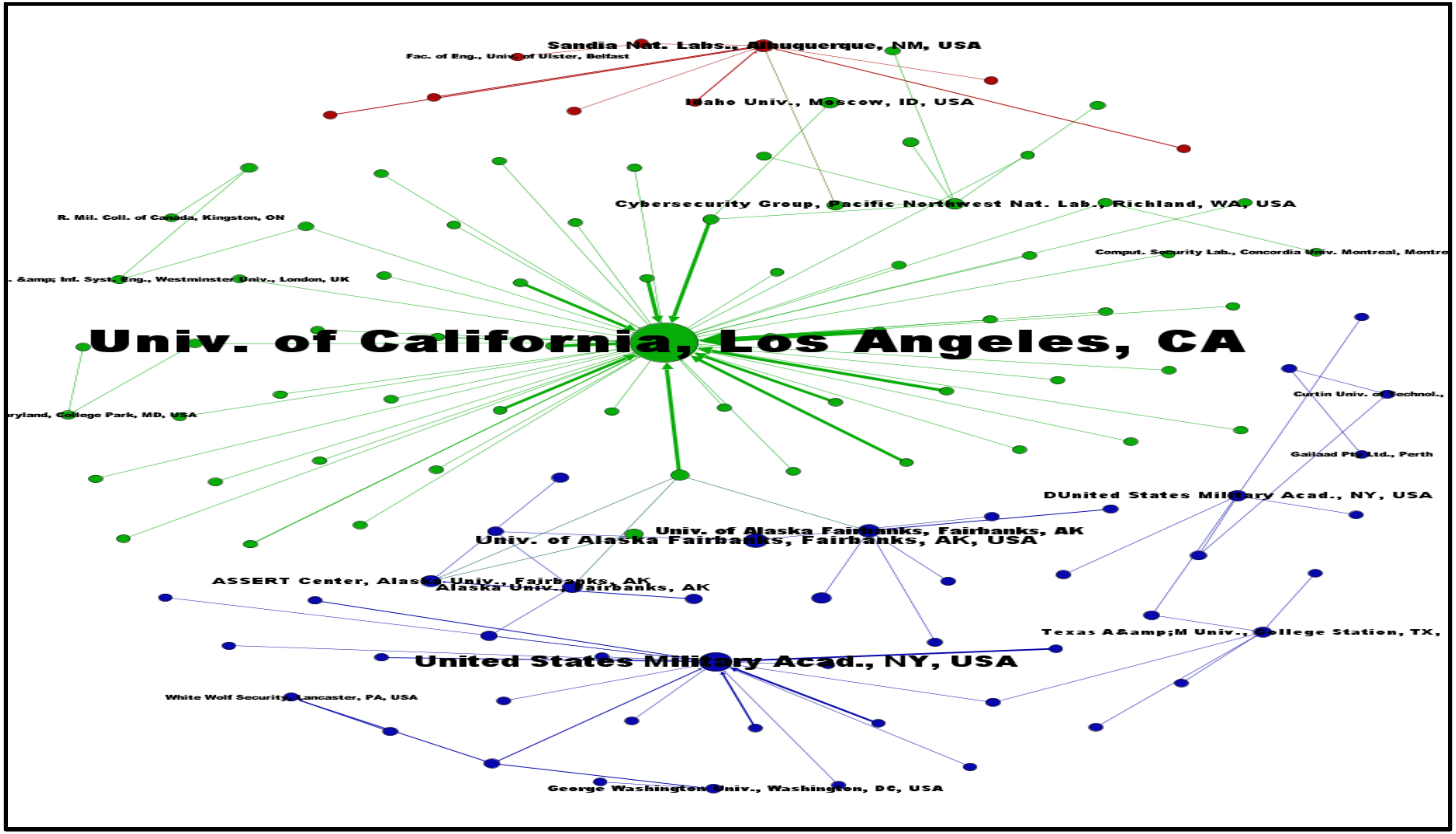

Figure 28: Cross-institutional Collaboration Network (MST)

RQ3: How has the network evolved?

RQ3.1: Has the community been able to absorb new members?

The ability of the IEEE and ACM community to attract novice researchers, whether doctoral students, new faculty members, or researchers who collaborate with other researchers on IS security, is a sign of an emergent IEEE and ACM network research community. *Figure 29* below presents a pictorial analysis of the *number of authors* who publish in IEEE and ACM, *the number of new authors* who publish in IEEE and ACM network for the very first time, and *the number of papers* accepted or published each year. Overall, the plotted diagram gives an indication of the trend analysis of the IEEE and ACM network for 1999 to 2013. The average number of papers accepted for publishing for that period is about 608. The graph for *Figure 29* illustrates a steady increase in the number of papers published between 1999 and 2010. There was a sharp increase in the number of papers published between 2008 and 2010 where it peaked at 1307. This was succeeded by a seemingly moderate decline for both 2010 and 2013. During 1999 and 2013 both the number of authors and new authors are directly proportional to the number of papers published. Hence, an increase in the number of new authors results in an increase in the number of papers accepted. The inverse is also true, when the number of papers accepted declines; we also see a reduction in the number of new authors.

The graph below (*Figure 29*) also shows that throughout the entire fourteen years of both programs, the number of new authors and the total number of authors were almost neck and neck. This near identical number explains the fact that researchers may have considered the maturity of the programs and as such has high confidence in collaborating with members of the research community.

Overall, it is difficult to pin point the factors that would lead to the dramatic decline in the number of new authors and other collaborating authors between 2010 and 2013 or the sustained increase in the number of authors from 1999 and 2011.

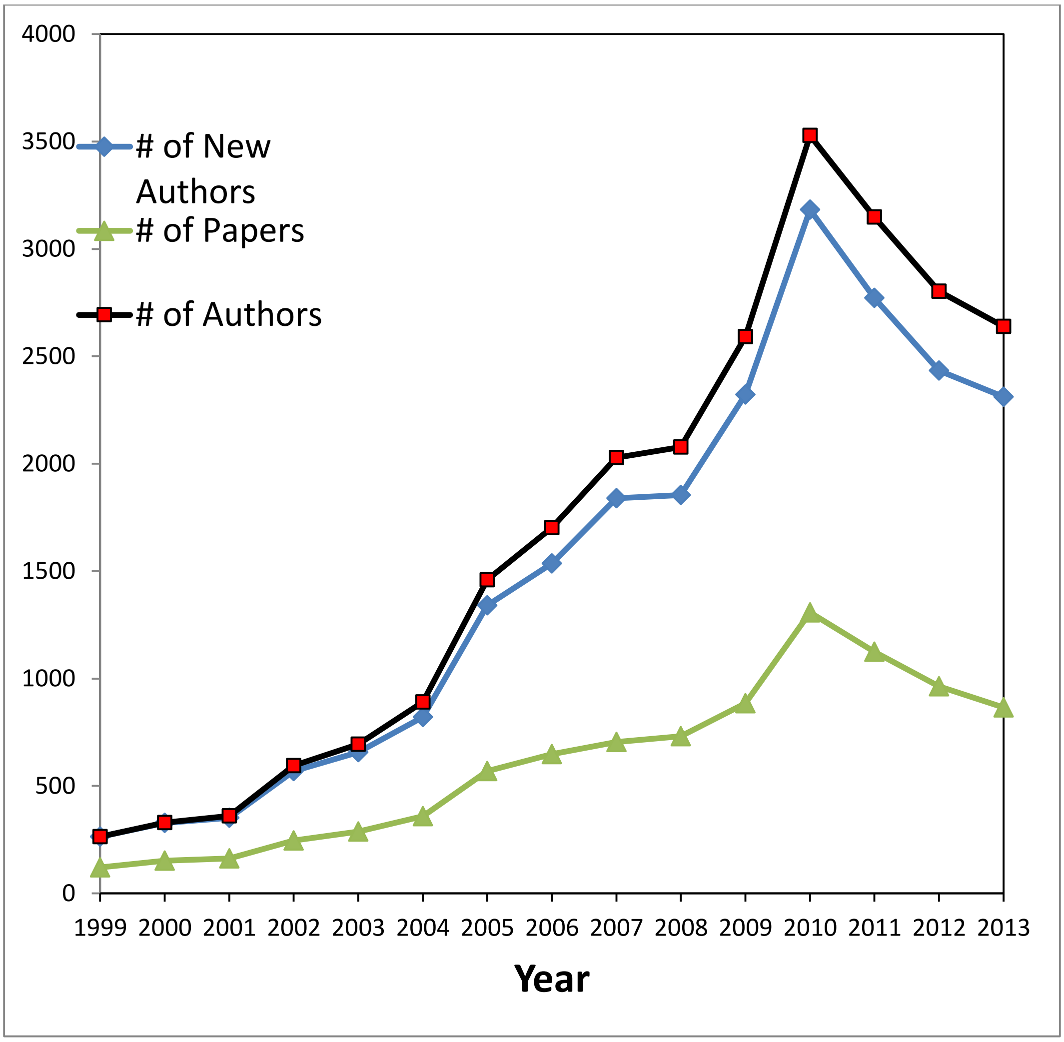

Figure 29: Changes in the Number of Papers, Number of Authors, and Number of new Authors over Time

RQ3.2: How has the structure of the network changed over time?

The structural characteristics of the IEEE and ACM network have evolved over the past fourteen years. Overall, new members have joined the network; others may have opted out, while existing members who had not previously collaborated have now begun to coauthor papers on a more regular basis. Moreover, new research topics may cause the emergence of new research groups and factions within the network.

To analyze the structure of the IEEE and ACM network, we divided it into three main temporal boundaries of five years each, namely *1999 -2004*, *1999 - 2008* and, *1999 - 2013*. Two authors are connected temporally, if they published papers together within that time window mentioned above. Figures *Figure* 30, *Figure* 31 and Figure 32 illustrate the MST structure of the IEEE and ACM network for the three time periods. *Figure 30* describes a loosely connected network with about three core groups of authors within the network. By 2009, the network structure had swelled to more than eight well-defined groups and about fifteen loosely scattered groups. The network structure continued to develop even further after 2009. In fact, the number of well-defined structures had more than quadrupled in 2013.

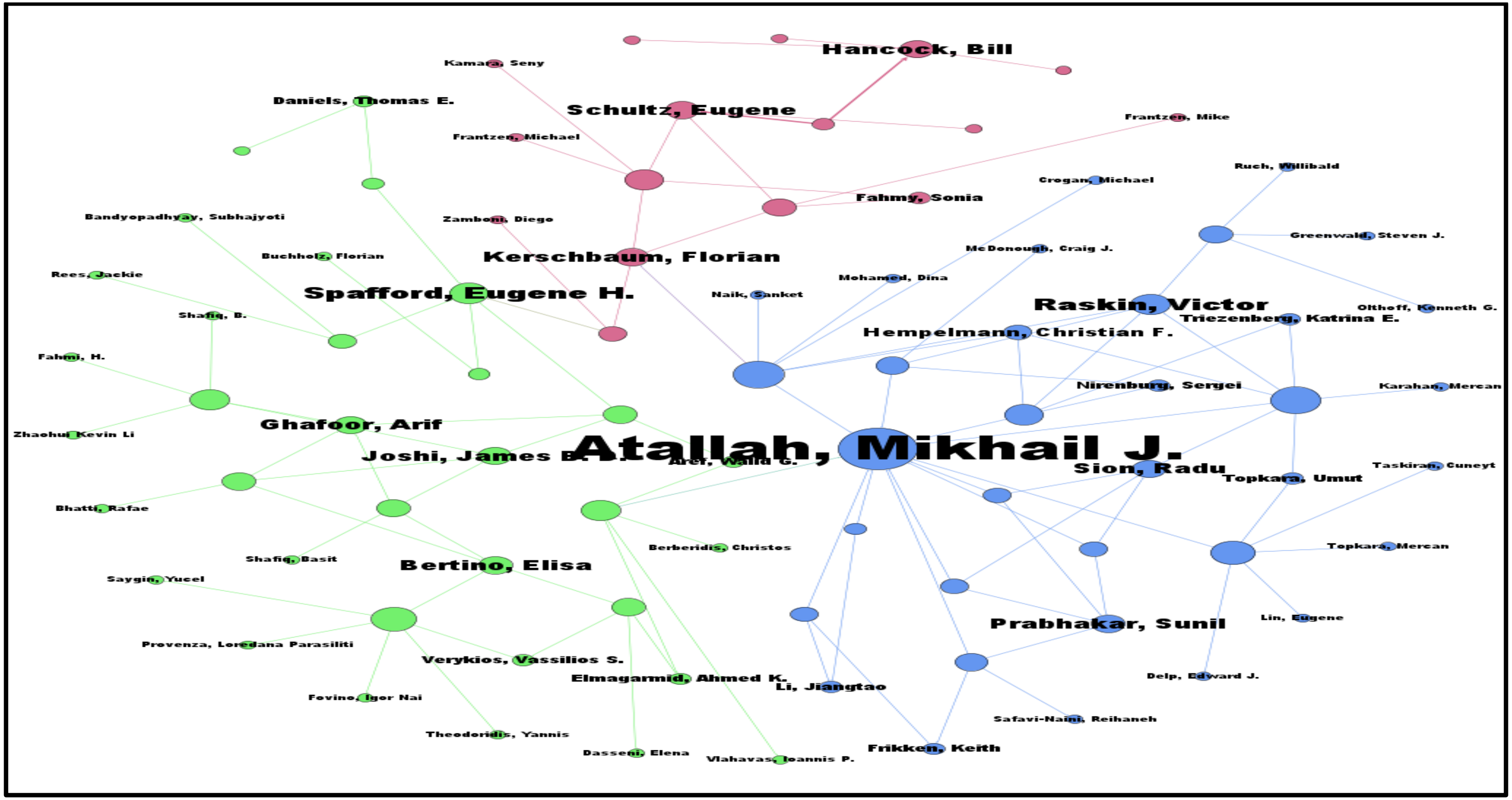

Figure 30: Structure of Giant Component of Network for 1999-2004

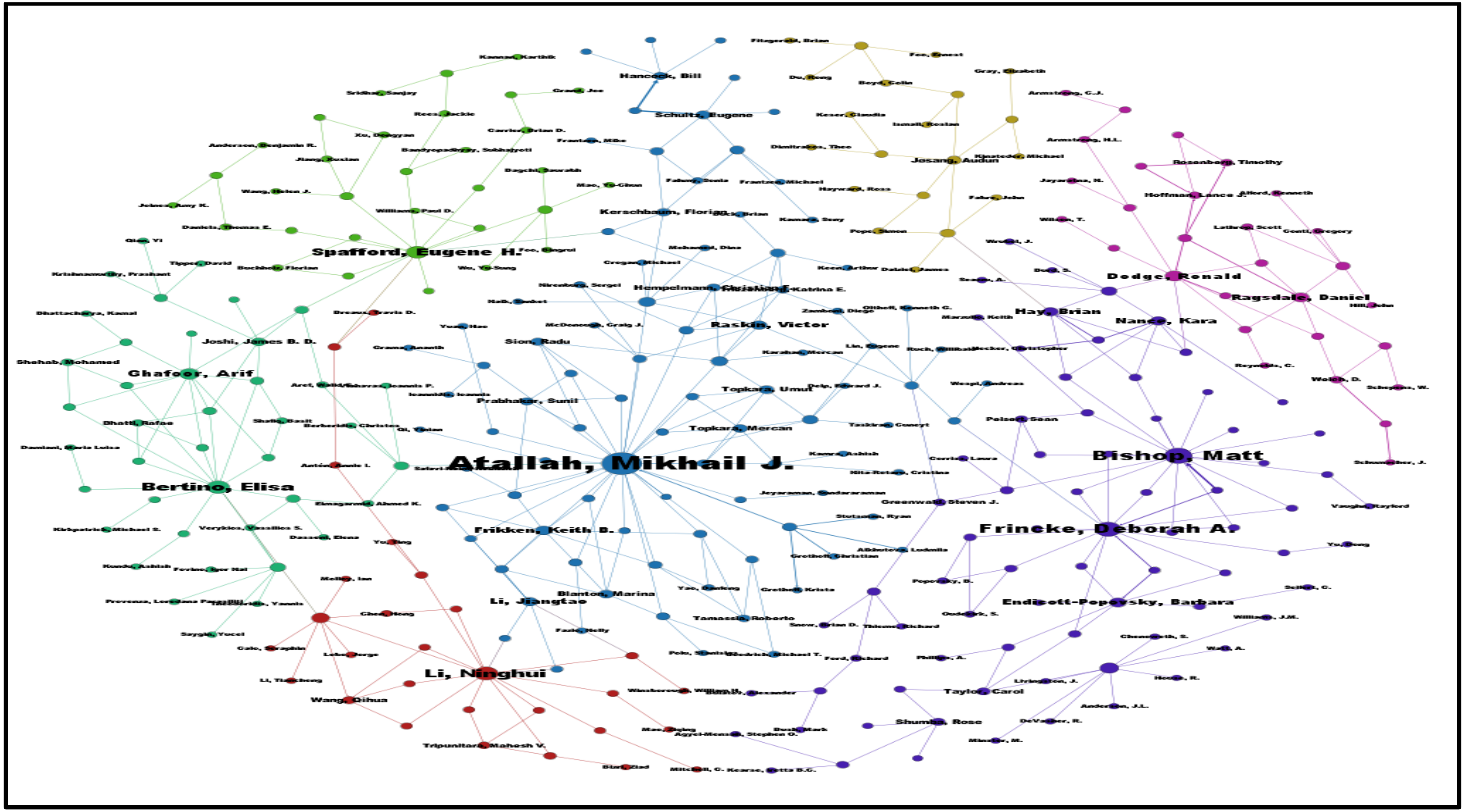

Figure 31: Structure of Giant Component of Network for 1999 - 2009

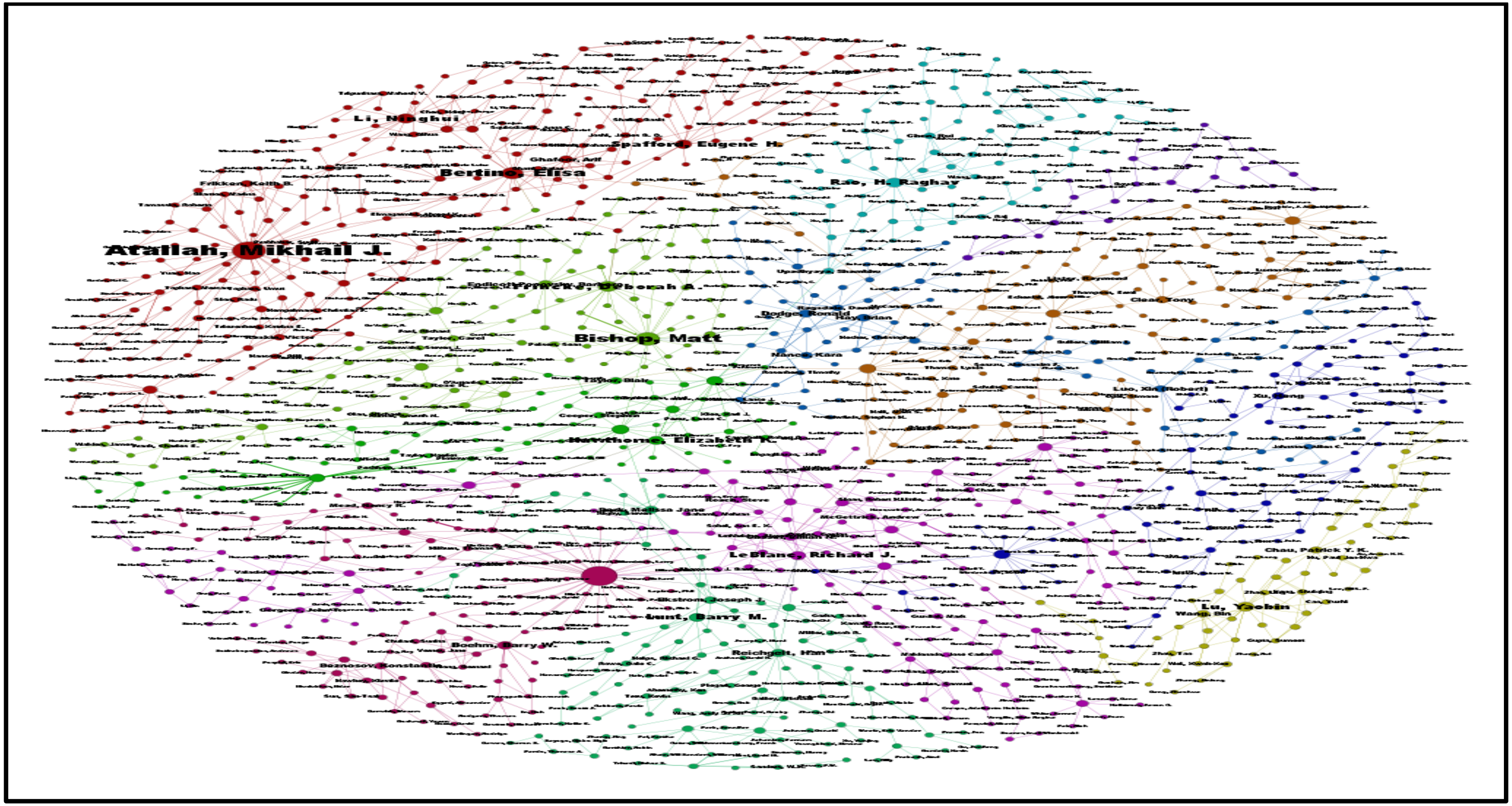

Figure 32: Structure of Giant Component of Network for 1999 - 2013

RQ3.3: Are there new generations and stars?

Table 19 and *Table 20* provide a listing of the most productive authors and institutions from 1999 to 2013. Although the authors have been collaborating on a significant basis, the same cannot be said for cross-institutional collaborations whether it is between regional institutions within the USA or international institutions. Overall, 8 of the top 10 most productive institutions located in the United States of America. Of those institutions, two of the top five are USA military based. Collaborations between authors of different institutions in the network grew with time. Table 19 shows the current star authors in the network. For a more accurate analysis, as a longitudinal study, it would be interesting to follow these authors for another ten to fifteen years to observe whether the group dynamics and structure have changed over time. It can also be observed from *Table 19* the star author for the network does not hail from the most productive institution in the network.

Table 19: Top Ten Productive Authors from 1999 - 2013

Author	**# of Papers**
1. Mikhail J. Atallah	39
2. Matt Bishop	28
3. Elisa Bertino	26
4. Mario Piattini	21
5. Raghav Rao	17
5. Ninghui Li	17
7. Eugene H. Spafford	15
8. Xiang-Yang Wang	14
8. Elizabeth Hawthorne	14
10. Brian Hay	13

Table 20: Top Ten Productive Institutions from 1999 - 2013

Institution	# of Papers
1.University of California, CA, USA	70
2. Carnegie Mellon University, PA, USA	30
3. US Military Academy, West Point, NY, USA	26
4. Iowa State Univ., Ames, IA, USA	23
5. Georgia Institute of Technology, GA, USA	20
6. Naval Postgraduate School, CA, USA	19
7. Sandia National Labs., Albuquerque, NM, USA	12
7. University of Southern California, CA, USA	12
9. California State Polytechnic University, CA, USA	11
10. North Carolina State University, North Carolina, USA	9

RQ3.4: How have collaboration patterns changed over time?

In this study, we categorize the percentage of papers with multiple authors as the *rate of author collaboration*. The percentage of papers that are written by authors from different institutions is regarded as the *rate of cross institutional collaboration*. *Figure 33* below presents the collaboration rates over time for the network. Clearly, the collaboration rate among authors occurs more frequently and is increasing more steadily and at a steeper slope than cross institutional collaboration. Though cross institutional collaborations experienced dramatic increases from 2002 to 2007, there has been relatively moderate but steady decreases since 2009.

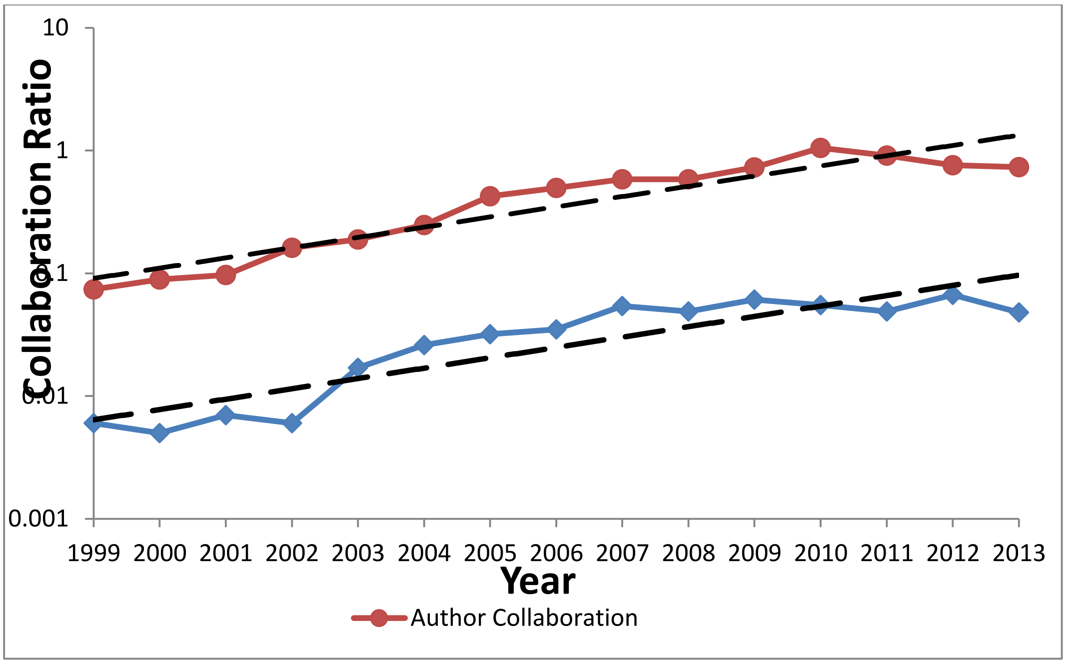

Figure 33: The Changes in Author Collaboration Rate and Institution Collaboration Rate Over Time

Section 4.2.5: Conclusions and Future Directions

Overall, our results revealed a weakly connected IEEE and ACM network. Many of the researchers have multiple collaborative papers. Our results underscore the existence of star researchers who either collaborate with other researchers or produce research papers individually. Further, we see a moderate increase in research collaboration at both institutional and cross-institutional levels. Above all, the trend indicates an increase in collaboration among researchers, which bodes well for the advancement, and stability of the IEEE and ACM communities.

We saw an increase in the number of individual, institutional collaboration and cross-institutional collaboration of publishers in both networks. However, many of these top institutions and publishers are from the United States of America … the international presence is lacking. In other to further enhance the identity of the IEEE and ACM network, a greater international flavor must be encouraged at all levels.

Finally, a closer investigation of the results shows that the most prolific institutions are the government agencies. This is can be clearly understood as the government, not the academic institutions, have a more vested interest in security related issues. A further examination shows that the top publishers are also from these government agencies. It is imperative that for the continued strength of the IEEE and ACM networks, a broader cross-section of the community must be incorporated into the system.

CHAPTER 5: CONCLUSIONS, CONTRIBUTIONS, AND FUTURE DIRECTIONS

5.1 Conclusion

In this thesis we presented two frameworks (a study and two cases) for analyzing the structure, topology and evolution of a traditional (static) social network as well as a collaborative and dynamic social network. Our study uses graph theory and a variety of social network analysis and visualization tools to analyze, solve and observe the structure of the static networks and an evolutionary point of view. In the first study we kept track of links and nodes of the governor's social network on Twitter in order to determine their network's topological features while also observing the elemental flow through the vertices.

We studied and analyzed when nodes and edges were added or deleted from the graphs in Case I and Case II of the co-authorship collaborative network to show how the structure and size of these graphs evolve on a temporal basis. In particular, we structurally fragmented the network to observe the prevalence of stars within the network and what would happen if those stars were to leave the network. Overall, the networks show characteristics that are pervasive in online social network features.

Specifically, in the first study, the focus was on using static social network analysis to determine how sitting United States governors utilize online social networks (Twitter) to distribute information to its citizens. We underscored many key conceptual theories of static SNA including structural balance, transitivity, reciprocity, social cohesion, influence, dominance, conformity and social role. Our analysis provided empirical evidence to support that transitivity and reciprocity occur with a high level of interaction and that there is very little dominance in the structure of the network. The result from the study supports other research studies on online social networks. Our research provides further confirmation to support structural balance (imbalance) of geographical homophily since the majority of the friends and followers of the governors are from their home State. Furthermore, the results suggest a network imbalance because most of the followers of the governors are located in the metropolitan areas resulting in the isolation of the dissemination of resources and services to citizens.

In Case I, dynamic SNA graph theory was used to analyze research collaborations of Information Assurance Education (IAE) researchers in an Information Systems Security Education domain. The primary emphasis of this case study was on analytic principles and concepts and the use of graph theory to represent the evolution of the collaborative network data. Using the graph theories, we utilized SNA analytic tools to help make predictions about their principal network structure. By following the dynamics of the collaborative network and the year the nodes were added allowed us to track the evolution of the social network. The study also allowed us to investigate various known characteristic of evolving social networks including centrality measures, degree, popularity, prestige, and so on. Our results seem to indicate that many of these characteristics are time dependent and that those properties evolve as the network expands by the addition of new authors or contracts by the deletion of existing author.

Case II is actually an extension of Case I. The case discussed many of the primary concepts of interpreting patterns of social ties among individuals in a dynamic network community. We examined the patterns of interactions among members of the network group to determine the structure of the network as well as the existence of cohesive sub-groups within the main network community. In addition, we analyzed the network to ascertain which central figure(s) play(s) key roles in the social network community. The study re-emphasized that not only is the structural prestige of a person in the social network a clear indication of the stability of the network but also a signal of the importance of social ties in the diffusion of information throughout the social system. Finally, we examined the social network to observe underlying correlating factors that have influenced the structure of the network with time.

Overall, the results for Case I and II revealed a weakly connected network. Many of the researchers in the network have multiple collaborative papers which suggests the confidence and social identity in the community. The results underscored the existence of star researchers who either collaborate with other researchers or produce research papers individually. Further, we saw a moderate increase in research collaboration at both institutional and cross-institutional levels. Above all, the trend indicates an increase in collaboration among researchers, which bodes well for the advancement, and stability of the communities.

We saw an increase in the number of individual, institutional collaboration and cross-institutional collaboration of publishers in both cases' networks. However, many of these top institutions and publishers are from the United States of America.

Finally, the true novelty and contribution of the dissertation is the fact that we were able to apply formal methods to a new area, that is, studying cybersecurity education and state government interaction, which has never done before.

5.2 Contributions

Whereas static and dynamic social network analysis symbolizes a range of theories relating types of observable social spaces and their relation to individual and group behavior and motivated by a structural intuition based on ties linking social actors, the practical use in real-world situations cannot be understated. These emergent structural properties of relational ties between actors are channels for transfer or "flow" of resources. For example, in the first study, we observed a high degree of geographic homophily in terms of the governors and their friends and followers which suggests that the majority of the governors' followers from their home state. Further, the analysis of the study highlights a disparity in the location of the governors' followers as most are located in the metropolitan areas. Since Twitter and many of the social network web services are highly pervasive, absolutely economical to use, and most importantly provide a practically almost instantaneous and unique method of communicating with friends and followers, the governors should do a better job in adding utility to their social network sites, especially their Twitter accounts so that their citizens in any part of their State could use this medium to obtain state updates. We can also apply the results of the concepts learnt to boost services and resources that may be inadequate in these remote areas.

It is absolutely clear that some of the concepts we investigated from the first study can enhance the understanding and the social structure of other network communities. This study can help us detect community structures that are real but hidden in many ways. For example, it can provide the framework to study metabolic networks and also the impetus to postulate other types of networks such as power grids, food webs and technological networks.

The two dynamic social network analysis case studies provide empirical data and graphical analysis of how social networks evolve over time, driven by the shared activities and affiliations of their members or by similarity of individuals' attributes. Those events that assist in the extension the network size and links with time (or vice versa) are critical events that characterize the evolution of the network community. We can use these events to our advantage (or disadvantage) if we are able to identify these readily and make to adjustments appropriately and where necessary. For instance, an understanding of the cases can be applied to better understand how quickly can diseases such as HIV and STDs spread through a population and which individuals to treat to slow down its spread. In addition, the methods used in this study can be applied to counter terrorism to determine the key figures in terror cells so that you can limit or thwart the threat of terrorism. The same can be applied to drug cells and the illegal drug trade.

Further, in the two cases we saw an increase in the number of individual, institutional collaboration and cross-institutional collaboration of publishers in IEEE and ACM networks. However, many of these top institutions and publishers are from the United States of America … the international presence is severely lacking. In other to further enhance the identity of the IEEE and ACM network, a greater international flavor must be encouraged at all levels.

Further, the most prolific institutions are the government agencies. This is can be clearly understood as the government, not the academic institutions, have a more vested interest in security related issues. Moreover, the top publishers are also from these government agencies. It is imperative that for the continued strength of the IEEE and ACM networks, a broader cross-section of the community must be incorporated into the system.

5.3 Limitations

The main assumption of the first study is that the nodes and links created throughout the social network are not stochastic in nature. We believed that the users are linked to other users by virtue of their geographical homophily. Establishing the structure and the dynamics of the homophily graph is an open problem, the solution to which will enable us to understand how, why and what aspects of homophily have on the system. For instance, we know that most governor's followers are from their state and reside in the metro areas. Are these people following the governors because they have or need the resources provided? Are they following because of the contents of the information that flow through the network? We believe that those factors could ultimately affect the topology of our graphs as well as our data analysis results.

In our model of co-authorship collaborations, we presented data at various temporal instances (from 1999 to 2013) with different time categories (5 year intervals). Although we have made several important approximations and deductions in explaining the topological structure of the network communities, we have on the other hand sacrificed certain network features for an analytical solution. For example, we neglected in our modeling effort the potential effect of age and the impact an author's retirement age may have on many of the social network analysis measurements such as closeness, prestige, betweenness, and so on.

A more detailed modeling of the co-authorship network would involve the construction of bipartite graphs [88], in which we directly simulate the publishing of papers by several co-authors, which are all connected to each other. In such a framework one can simultaneously study the evolution of the co-authorship network, in which nodes are linked by joint publications and the publication network, in which nodes are papers linked by joint authors [135]. This such bipartite network could easier predict the network's dynamics and topology, however, this is beyond the objective of this study.

5.4 Future Directions

The main applicative perspective of using our social network analysis approach in the first study is to assist in the optimization of the flow of resources within the elected government community's social network. On a broader perspective, we could utilize this same approach on a national level to improve the performance of resource flow for citizens on a regional and national basis.

Twitter is a pervasive form of technology that allows us to engage in a conversation and communication. Due to its ease of use and its low network resource utilization, it is rapidly being adopted in the community. Citizens seem to follow their governors primarily because they are interested in the subjects and contents of their tweets. Since in our research we were not determining why the governors are following federal government agencies, for future work, it would be worthwhile to know why state governors are following certain individuals and state agencies. Knowing why the governors are following the federal agencies could be beneficial to the federal government when disseminating any kind of resources for its citizens, for example in times of a natural disaster or the spread of diseases; economic resources, such as areas that need the most help could receive it adequately and in a timely manner.

A more detailed modeling of the co-authorship network would involve the construction of bipartite graphs [88], in which we directly simulate the publishing of papers by several co-authors, which are all connected to each other. In such a framework one can simultaneously study the evolution of the co-authorship network, in which nodes are linked by joint publications and the publication network, in which nodes are papers linked by joint authors [135]. This such bipartite network could easier predict the network's dynamics and topology, however, this is beyond the objective of this study.

Although we have seen an increase in the number of individual, institutional collaboration and cross-institutional collaboration of publishers in the area on information assurance education and security however, many of these top institutions and publishers are from the United States of America. In other to further enhance the identity of the information assurance education and security education a greater international flavor must be encouraged at all levels. One way to increase participation, is to alternate conferences in national and international venues. In addition, cross training of staff in the discipline should be fostered so that participation can be improved.

In summary, much of the graph model analysis presented in this dissertation was done using social networks analysis tools and depicted using common social networks analysis visualization tools. Although, this helps in determining an acceptable solution to the problem, there are however some key network properties that may go amiss. It could be worthwhile if we can extend the analysis using semi-manual or mathematics formulas to determine whether we will have similar results. Such method could enable us to uncover other subtle network properties that may be hidden in the network structure.

Bibliography

[1] M. E. J. Newman, “The Structure and Function of Complex Networks,” *SIAM Review*, vol. 45, no. 2. pp. 167–256, 2003.

[2] D. Knoke and S. Yang, *Social Network Analysis*, vol. 8, no. 3. 2008.

[3] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, vol. 8, no. 1. Cambridge University Press, 1994.

[4] C. Prell, *Social Network Analysis: History, Theory and Methodology*, vol. 2011. SAGE, 2011.

[5] W. De Nooy, A. Mrvar, and V. Batagelj, *Exploratory social network analysis with Pajek*, 2nd ed., vol. 40, no. 3. Cambridge University Press, 2011.

[6] A. Bavelas, “Communication patterns in task-oriented groups,” *J. Acoust. Soc. Am.*, vol. 22, no. 6, pp. 725–730, 1950.

[7] S. Milgram, “The small world problem,” *Psychol. Today*, vol. 1, no. 1, pp. 60–67, 1967.

[8] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks.,” *Nature*, vol. 393, no. 6684, pp. 440–2, 1998.

[9] P. Erdös and A. Rényi, “On random graphs,” *Publ Math Debrecen*, vol. 6, no. 290–297, pp. 290–297, 1959.

[10] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, “Measurement and analysis of online social networks,” *Proc. 7th ACM SIGCOMM Conf. Internet Meas. - IMC ’07*, vol. 40, no. 6, p. 29, 2007.

[11] O. Frank and D. Strauss, “Markov Graphs,” *J. Am. Stat. Assoc.*, vol. 81, no. 395, pp. 832–842, 1986.

[12] P. Pattison and S. Wasserman, “Logit models and logistic regressions for social networks: II. Multivariate relations.,” *Br. J. Math. Stat. Psychol.*, vol. 52 (Pt 2), no. 2, pp. 169–193, 1999.

[13] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 29, no. 4, pp. 251–262, 1999.

[14] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science (80-.).*, vol. 286, no. 5439, p. 11, 1999.

[15] L. A. Adamic, O. Buyukkokten, and E. Adar, "A social network caught in the Web," *First Monday*, vol. 8, no. 6, pp. 1–22, 2003.

[16] M. Cha, A. Mislove, and K. P. Gummadi, "A measurement-driven analysis of information propagation in the flickr social network," *Proc. 18th Int. Conf. World wide web - WWW '09*, p. 721, 2009.

[17] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and a.-L. Barabasi, "The large-scale organization of metabolic networks," *Nature*, vol. 407, pp. 651–654, Oct. 2000.

[18] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, "An introduction to exponential random graph (p) models for social networks," *Soc. Networks*, vol. 29, no. 2, pp. 173–191, 2007.

[19] L. C. Freeman, *The development of social network analysis*, vol. 27, no. 3. 2004.

[20] E. Otte and R. Rousseau, "Social network analysis: a powerful strategy, also for the information sciences," *J. Inf. Sci.*, vol. 28, no. 6, pp. 441–453, 2002.

[21] M. Á. Serrano and M. Boguná, "Clustering in complex networks. II. Percolation properties," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 74, no. 5, pp. 1–8, 2006.

[22] R. a Hanneman, "Introduction to Social Network Methods," *Network*, vol. 2, no. 3, pp. 292–1990, 1998.

[23] J. Douceur, "The sybil attack," *Peer-to-peer Syst.*, pp. 251–260, 2002.

[24] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender Systems with Social Regularization Categories and Subject Descriptors," *Proc. fourth ACM*

Int. Conf. Web search data Min., pp. 287–296, 2011.

[25] A. Mislove, A. Post, P. Druschel, and K. P. Gummadi, "Ostra: Leveraging Trust to Thwart Unwanted Communication," *NSDI '08 Proc. 5th USENIX Symp. Networked Syst. Des. Implement.*, no. i, pp. 15–30, 2008.

[26] N. Z. Gong, W. Xu, L. Huang, P. Mittal, E. Stefanov, V. Sekar, and D. Song, "Evolution of Social-Attribute Networks: Measurements, Modeling, and Implications using Google+," *Internet Meas. Conf.*, pp. 131–144, 2012.

[27] C. Thovex and F. Trichet LINA, "Static and Semantic Social Networks Analysis: Towards a Multidimensional Convergent Model."

[28] C. Thovex and F. Trichet, "Static, Dynamic and Semantic Dimensions: Towards a Multidisciplinary Approach of Social Networks Analysis," *Knowl. Sci. Eng. Manag.*, vol. 6291, pp. 567–572, 2010.

[29] V. Latora and M. Marchiori, "Efficient behavior of small-world networks.," *Phys. Rev. Lett.*, vol. 87, no. 19, p. 198701, 2001.

[30] D. Greene, D. Doyle, and P. Cunningham, "Tracking the evolution of communities in dynamic social networks," in *Proceedings - 2010 International Conference on Advances in Social Network Analysis and Mining, ASONAM 2010*, 2010, pp. 176–183.

[31] Y. Lin, "FacetNet : A Framework for analyzing communities and their evolutions in dynamic networks," *Soc. Networks*, pp. 685–694, 2008.

[32] G. Palla, A.-L. Barabási, and T. Vicsek, "Quantifying social group evolution," *Nature*, vol. 446, no. 7136, pp. 664–667, 2007.

[33] B. Wu, Q. Ye, S. Yang, and B. Wang, "Group CRM: a new telecom CRM framework from social network perspective," *Proceeding 1st ACM Int. Work. Complex networks meet Inf. Knowl. Manag. - CNIKM '09*, pp. 3–10, 2009.

[34] D. I. Rubenstein, S. R. Sundaresan, I. R. Fischhoff, C. Tantipathananandh, and T. Y. Berger-wolf, "Similar but Different : Dynamic Social Network Analysis Highlights Fundamental Differences between the Fission-Fusion Societies of Two Equid Species , the Onager and Grevy ' s Zebra," pp. 1–21, 2015.

[35] T. Y. Berger-Wolf, C. Tantipathananandh, and D. Kempe, “Dynamic Community Identification,” in *Link Mining: Models, Algorithms, and Applications*, P. S. Yu, J. Han, and C. Faloutsos, Eds. Springer New York, 2010, pp. 45–47.

[36] D. de Kool and J. van Wamelen, “Web 2.0: A New Basis for E-Government?,” in *2008 3rd International Conference on Information and Communication Technologies From Theory to Applications*, 2008, pp. 1–7.

[37] T. O’Reilly, “What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software,” *Design*, vol. 65, no. 65, pp. 17–37, 2007.

[38] F. Wigand, “Twitter takes wing in government: diffusion, roles, and management,” *11th Annu. Int. Digit. Gov.*, pp. 66–71, 2010.

[39] DoD, “The Use of Web 2.0 in the Department of Defense,” 2009.

[40] G. C. Wilshusen, “Challenges In Federal Agencies’ Use of Web 2.0 Technologies,” Washington,DC, 2010.

[41] I. Mergel, C. Schweik, and J. Fountain, “The Transformational Effect of Web 2 . 0 Technologies on Government,” 2009.

[42] H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter , a Social Network or a News Media? Categories and Subject Descriptors,” *Most*, vol. 112, no. 2, pp. 591–600, 2010.

[43] D. Boyd and S. Golder, “Tweet, tweet, retweet: Conversational aspects of retweeting on twitter,” in *System Sciences (HICSS), 2010*, 2010, pp. 1–10.

[44] D. Boyd, S. Golder, and G. Lotan, “Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter,” in *2010 43rd Hawaii International Conference on System Sciences*, 2010, vol. 0, no. 6, pp. 1–10.

[45] A. Lenhart, K. Purcell, A. Smith, and K. Zickuhr, “Social Media & Mobile Internet Use Among Teens and Young Adults,” *Pew Internet Am. Life Proj.*, vol. 5, no. 1, pp. 1–37, 2010.

[46] F. D. L. Wigand, “Twitter in Government: Building Relationships One Tweet at a Time,” *2010 Seventh Int. Conf. Inf. Technol. New Gener.*, pp. 563–567,

2010.

[47] M. Skoric, N. Poor, P. Achananuparp, E.-P. Lim, and J. Jiang, “Tweets and Votes: A Study of the 2011 Singapore General Election,” in *2012 45th Hawaii International Conference on System Sciences*, 2011, pp. 2583–2591.

[48] A. Chang and P. K. Kannan, “Leveraging Web 2 . 0 in Government E-Government / Technology Series,” 2008.

[49] D. Osimo, “Web 2 . 0 in Government: Why and How?,” *Inst. Prospect. Technol. Stud. IPTS JRC Eur. Comm.*, vol. 23358, 2008.

[50] A. Smith, “Government Online information,” *Methodology*, pp. 1–44, 2010.

[51] R. D. Waters and J. M. Williams, “Academic Paper Squawking , tweeting , cooing , and hooting : analyzing the communication patterns of government agencies on Twitter,” *J. Public Aff.*, vol. 11, no. 4, pp. 353–363, 2011.

[52] J. Golbeck, J. Grimes, and A. Rogers, “Twitter Use by the U . S . Congress,” *J. Am. Soc. Inf. Sci.*, vol. 61, no. 8, pp. 1612–1621, 2010.

[53] F. Kivran-swaine, P. Govindan, and M. Naaman, “The Impact of Network Structure on Breaking Ties in Online Social Networks: Unfollowing on Twitter,” *Hum. Factors*, no. 02, pp. 1–4, 2011.

[54] H. Zhang, R. Dantu, and J. Cangussu, “Quantifying Reciprocity in Social Networks,” in *2009 International Conference on Computational Science and Engineering*, 2009, pp. 1031–1035.

[55] S. Yardi, “Tweeting from the Town Square: Measuring Geographic Local Networks,” *Methods*, vol. 77, no. 3, pp. 194–201, 2010.

[56] A. Java, X. Song, T. Finin, and B. Tseng, “Why we twitter: understanding microblogging usage and communities,” *Network*, vol. 43, no. 1, pp. 56–65, 2007.

[57] R. Kumar, J. Novak, and A. Tomkins, “Structure and evolution of online social networks,” in *12th ACM SIGKDD international conference on Knowledge discovery and data mining KDD 06*, 2006, p. 611.

[58] H. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas, “Homophily in the Digital World: A LiveJournal Case Study,” *IEEE Internet Comput.*, vol. 14, no. 2, pp. 15–23, Mar. 2010.

[59] M. McPherson, L. Smith-Lovin, and J. M. Cook, “Birds of a Feather: Homophily in Social Networks,” *Annu. Rev. Sociol.*, vol. 27, no. 1, pp. 415–444, Aug. 2001.

[60] D. Zhao and M. B. Rosson, “How and Why People Twitter : The Role that Micro-blogging Plays in Informal Communication at Work,” *Hum. Factors*, pp. 243–252, 2009.

[61] S. Everton, “Tracking, Destabilizing, and Disrupting Dark Networks with Social Network Analysis,” *Dark Networks Course Man.*, 2010.

[62] M. S. Granovetter, “The Strength of Weak Ties,” *Am. J. Sociol.*, vol. 78, no. 6, pp. 1360–1380, 1973.

[63] G. R. Azarian*, The General Sociology of Harrison C. White: Chaos and order in networks*. Palgrave Macmillan, 2005.

[64] M. O. Jackson, “An Overview of Social Networks and Economic,” in *The handbook of social economics* , 2010, pp. 1–96.

[65] A. Barabasi, R. Albert, and H. Jeong, “Scale-free characteristics of random networks: the topology of the world-wide web,” *Phys. A Stat. Mech. its Appl.*, vol. 281, no. 1–4, pp. 69–77, 2000.

[66] J. Xu and M. Chau, “The Social Identity of IS: Analyzing the Collaboration Network of the ICIS Conferences(1980-2005) General Topics,” pp. 569–590, 2006.

[67] S. P. Borgatti, “Centrality and network flow,” *Soc. Networks*, vol. 27, no. 1, pp. 55–71, Jan. 2005.

[68] B. Priyopradono, D. Manongga, and W. Herry Utomo, “Spatial Social Network Analysis: Program Pengembangan Usaha Agribisnis Perdesaan (PUAP) or an Exertion Development Program in Supporting the Region Revitalization Development,” *Soc. Netw.*, vol. 02, no. 02, pp. 63–76, 2013.

[69] J. A. Rodríguez, "The March 11th Terrorist Network: In Its Weakness Lies Its Strength," *XXV Int. Sunbelt*, 2005.

[70] F. J. Acedo, C. Barroso, C. Casanueva, and J. L. Galán, "Co-authorship in management and organizational studies: An empirical and network analysis," *J. Manag. Stud.*, vol. 43, no. 5, pp. 957–983, 2006.

[71] G. Kossinets and D. J. Watts, "Empirical analysis of an evolving social network.," *Science*, vol. 311, no. 5757, pp. 88–90, 2006.

[72] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 69, no. 2 2, pp. 1–15, 2004.

[73] A. Abbasi and J. Altmann, "On the correlation between research performance and social network analysis measures applied to research collaboration networks," *... 44th Hawaii Int. Conf.*, p. 22, 2010.

[74] K. G. Provan, M. a. Veazie, L. K. Staten, and N. I. Teufel-Shone, "The use of network analysis to strengthen community partnerships," *Public Adm. Rev.*, vol. 65, no. 5, pp. 603–612, 2005.

[75] D. Krackhardt, "The Strength of Strong Ties: The Importance of Philos in Organizations," in *Networks and organizations : structure, form, and action*, N. Nohria and R. G. Eccles, Eds. Harvard Business School Press, 1992, p. xvi, 544 p.

[76] M. E. J. Newman, "Models of the Small World," *J. Stat. Phys.*, vol. 101, no. 3, pp. 819–841, 2000.

[77] a Abbassi, H. Liaquat, and L. Leydesdorff, "Betweeness centrality as a driver of preferential attachment in the evolution of research collaboration networks," *J. Informetr.*, vol. 6, no. 3, pp. 403–412, 2012.

[78] X. Liu, J. Bollen, M. L. Nelson, and H. Van De Sompel, "Co-authorship networks in the digital library research community," *Inf. Process. Manag.*, vol. 41, no. 6, pp. 1462–1480, 2005.

[79] M. E. J. Newman, "A Measure of Betweenness Based on Random Walks," *arXiv:cond-mat*, no. 0309045, pp. 1–15, 2003.

[80] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, “Classes of small-world networks,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 21, pp. 11149–11152, 2000.

[81] J. Moody, “The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999,” *Am. Sociol. Rev.*, vol. 69, no. 2, pp. 213–238, 2004.

[82] N. E. FRIEDKIN, “Structural Cohesion and Equivalence Explanations of Social Homogeneity,” *Sociological Methods & Research*, vol. 12, no. 3. pp. 235–261, 1984.

[83] Z.-Z. Wang and J. J. H. Zhu, “Homophily versus preferential attachment: Evolutionary mechanisms of scientific collaboration networks,” *Int. J. Mod. Phys. C*, vol. 25, no. 05, p. 1440014, 2014.

[84] M. McPherson, L. Smith-Lovin, and J. M. Cook, “Birds of a Feather: Homophily in Social Networks,” *Annual Review of Sociology*, vol. 27, no. 1. pp. 415–444, 2001.

[85] H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter , a Social Network or a News Media?,” *Int. World Wide Web Conf. Comm.*, pp. 1–10, 2010.

[86] S. A. Golder and S. Yardi, “Structural Predictors of Tie Formation in Twitter: Transitivity and Mutuality,” in *Methods*, 2010, pp. 88–95.

[87] P. Macdonald, E. Almaas, and a. L. Barabási, “Minimum spanning trees of weighted scale-free networks,” *EPL (Europhysics Lett.*, vol. 72, no. i, p. 308, 2005.

[88] M. E. J. Newman, D. J. Watts, and S. H. Strogatz, “Random graph models of social networks.,” *Pnas*, vol. 99 Suppl 1, pp. 2566–72, 2002.

[89] M. Pósfai, G. Musella, M. Martino, R. Sinatra Acknowledgements, S. Morrison, A. Husseini, and P. Hoevel, “Chapter 5: The Barabási-Albert Model,” in *Network Science*, 2015.

[90] A. Clauset, C. R. Shalizi, and M. E. J. Newman, “Power-Law Distributions in Empirical Data,” *SIAM Rev.*, vol. 51, no. 4, p. 661, 2009.

[91] T. Schank and D. Wagner, “Approximating Clustering Coefficient and Transitivity,” *J. Graph Algorithms Appl.*, vol. 9, no. 2, pp. 265–275, 2005.

[92] A. Barrat and M. Weigt, “On the properties of small-world network models,” *Eur. Phys. J. B*, vol. 13, no. 3, pp. 547–560, 2000.

[93] J. a. Shuler, P. T. Jaeger, and J. C. Bertot, “Implications of harmonizing the future of the federal depository library program within e-government principles and policies,” *Gov. Inf. Q.*, vol. 27, no. 1, pp. 9–16, Jan. 2010.

[94] J. C. Bertot, P. T. Jaeger, and J. M. Grimes, “Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies,” *Gov. Inf. Q.*, vol. 27, no. 3, pp. 264–271, Jul. 2010.

[95] V. Cheterian, “GCSP Policy Paper n°11,” *Geneva Cent. Secur. Policy*, no. February, 2011.

[96] M. Ottaway and A. Hamzawy, “Protest Movements and Political Change in the Arab World,” *Carnegie Endow. Policy Outlook*, no. January 2011, pp. 1–14, 2011.

[97] P. T. Metaxas, E. Mustafaraj, and D. Gayo-Avello, “How (Not) To Predict Elections,” in *2011 IEEE Third Intl Conference on Privacy Security Risk and Trust and 2011 IEEE Third Intl Conference on Social Computing*, 2011, pp. 165–171.

[98] R. Balasubramanyan, B. R. Routledge, and N. A. Smith, “From tweets to polls: Linking text sentiment to public opinion time series,” in *Proceedings of the International AAAI Conference on Weblogs and Social Media*, 2010, no. May, pp. 122–129.

[99] J. Chung and E. Mustafaraj, “Can Collective Sentiment Expressed on Twitter Predict Political Elections?,” *Artif. Intell.*, vol. 2, pp. 1770–1771, 2010.

[100] S. Cohen and W. Eimicke, “The Future of E-Government: A Project of Potential Trends and Issues School of International and Public Affairs,” *New York*, vol. 00, no. C, pp. 1–10, 2003.

[101] C. Shirky, “Power Laws, Weblogs, and Inequality,” *Clay Shirky’s Writings About the Internet*, vol. 8, no. 16/03/2004. 2003.

[102] A.-L. Barabási and R. Albert, “Emergence of Scaling in Random Networks,” *Science*, vol. 286, no. 5439, pp. 1–11, 1999.

[103] A. Java, X. Song, T. Finin, and B. Tseng, “Why we twitter: understanding microblogging usage and communities,” *Network*, vol. 43, no. 1, pp. 56–65, 2007.

[104] M. J. Welch, U. Schonfeld, D. He, and J. Cho, “Topical Semantics of Twitter Links,” *Time*, vol. 30, no. 2, pp. 327–336, 2011.

[105] L. C. Pérez, S. Cooper, E. K. Hawthorne, S. Wetzel, J. Brynielsson, K. Klee, J. Impagliazzo, N. Pohlmann, M. Leary, A. Philips, B. Taylor, and B. Hall, “Information Assurance Education in Two- and Four-Year Institutions,” in *16th Annual conference reports on Innovation and technoogy in computer science education-working group reports*, 2011, pp. 39–53.

[106] V. Padman and N. Memon, “Design of A Virtual Laboratory for Information Assurance Education and Research,” *Work. Inf. Assur. Secur.*, vol. 1, 2002.

[107] W. Schepens and J. R. Surdu, “The Cyber Defense Exercise: An evaluation of the effectiveness of information assurance education,” *J. Inf. Secur.*, vol. 1, no. 2, 2002.

[108] D. Schweitzer, J. Humphries, and L. Baird, “Meeting the criteria for a Center of Academic Excellence (CAE) in information assurance education,” *J. Comput. Sci. Coll.*, vol. 22, no. 151–160, 2006.

[109] G. DeSanctis, “The Social Life of Information Systems Research: A Response to Benbasat and Zmud’s Call for Returning to the IT Artifact,” *J. Assoc. Inf. Syst.*, vol. 4, no. 7, pp. 360–376, 2003.

[110] S. Cooper, W. Lafayette, C. Nickell, F. Meade, V. Piotrowski, B. Oldfield, W. Dc, A. Abdallah, M. Bishop, B. Caelli, L. Hoffman, L. C. Pérez, C. Pfleeger, R. Raines, C. Schou, and J. Brynielsson, “An Exploration of the Current State of Information Assurance Education,” vol. 41, no. 4, 2009.

[111] S. Azadegan, M. Lavine, M. O’Leary, a. Wijesinha, and M. Zimand, “An undergraduate track in computer security,” *Proc. 8th Annu. Conf. Innov. Technol. Comput. Sci. Educ. - ITiCSE ’03*, p. 207, 2003.

[112] M. O’Leary, S. Azadegan, and J. Lakhani, “Development of a Honeynet

Laboratory: a Case Study," *Seventh ACIS Int. Conf. Softw. Eng. Artif. Intell. Networking, Parallel/Distributed Comput.*, pp. 401–406, 2006.

[113] S. P. Borgatti, M. G. Everett, and L. C. Freeman, "Ucinet for Windows: Software for Social Network Analysis," *Harvard Anal. Technol.*, vol. 2006, p. SNA Analysis software, 2002.

[114] M. Crawley, *The R book*. Wiley, 2007.

[115] R. R Development Core Team, "R: A Language and Environment for Statistical Computing," *R Foundation for Statistical Computing*, vol. 1, no. 2.11.1. R Foundation for Statistical Computing, p. 409, 2011.

[116] D. Qian, M. D. Gross, G. Augenbroe, C. Eastman, and M. D. Gross, "Collaborative Design with NetDraw Collaborative Design with NetDraw," *Scenario*, vol. 99, 1999.

[117] H. D. Pannhorst and W. J. Schaffer, "NETDRAW, a NETOMAC® GUI for power engineering education," *Proc. Univ. Power Eng. Conf.*, p. 116, 2000.

[118] N. M. Tichy, M. L. Tushman, and C. Fombrun, "Analysis," *Acad. Manag. Rev.*, vol. 4, no. 4, pp. 507–519, 1979.

[119] N. Erman and L. Todorovski, "Mapping the E-Government Research with Social Network Analysis," in *Eighth International Conference EGOV 2009*, vol. 5693, M. A. Wimmer, H. J. Scholl, M. Janssen, and R. Traunmüller, Eds. Springer Verlag, 2009, pp. 13–25.

[120] E. Castilla, H. Hwang, E. Granovetter, and M. Granovetter, "Social Networks in Silicon Valley," in *The Silicon Valley Edge: A Habitat for Innovation and Entrepreneurship*, 2000, pp. 218–247.

[121] U. Brandes, J. Raab, and D. Wagner, "Exploratory Network Visualization: Simultaneous Display of Actor Status and Connections," *Public Policy*, vol. 2, no. 4, p. 28, 2001.

[122] B. Hogan, J. A. Carrasco, and B. Wellman, "Visualizing Personal Networks: Working with Participant-aided Sociograms," *Field methods*, vol. 19, no. 2, pp. 116–144, 2007.

[123] a. L. Barabási, H. Jeong, Z. Néda, E. Ravasz, a. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations," *Phys. A Stat. Mech. its Appl.*, vol. 311, no. 3–4, pp. 590–614, 2002.

[124] P. Crucitti, V. Latora, M. Marchiori, and A. Rapisarda, "Efficiency of scale-free networks: Error and attack tolerance," *Phys. A Stat. Mech. its Appl.*, vol. 320, pp. 622–642, 2003.

[125] R. Chase and S. Borgatti, "http://www.analytictech.com/keyplayer/kpsetup.exe," 2012. [Online]. Available: http://www.analytictech.com/.

[126] THE WHITE HOUSE, "Remarks by the President on Securing Our Nation's Cyber Infrastructure," 2009. [Online]. Available: https://www.whitehouse.gov/video/President-Obama-on-Cybersecurity#transcript.

[127] The Joint Task Force on Computing Curricula (Association for Computing Machinery IEEE-Computer Society), *Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science*. 2013.

[128] "National Centers of Academic Excellence - Cyber Defense," *National Security Agency*, 2015. [Online]. Available: https://www.nsa.gov/academia/ncae-cd/index.shtml.

[129] D. Frincke and M. Bishop, "Joining the security education community," *IEEE Security and Privacy*, vol. 2, no. 5. pp. 61–63, 2004.

[130] W. V. Maconachy, C. D. Schou, D. Ragsdale, and D. Welch, "A Model for Information Assurance : An Integrated Approach," in *Proceedings of the 2001 IEEE Workshop on Information Assurance and Security*, 2001, pp. 5–6.

[131] C. J. R. McCumber, "Information Systems Security: A Comprehensive Model," in *proceedings of the 14th National Computer Security Conference, October 1991*, 1991, no. 4011, pp. 1–12.

[132] M. Bishop, "What is computer security?," *Secur. Privacy, IEEE*, vol. 1, no. 1, pp. 67–69, 2003.

[133] C. E. Irvine, S. K. Chin, and D. Frincke, "Integrating security into the

curriculum," *Computer (Long. Beach. Calif).*, vol. 31, no. 12, pp. 25–30, 1998.

[134] M. Bishop, "Computer security education: training, scholarship, and research," *Computer (Long. Beach. Calif).*, vol. 35, no. 4, 2002.

[135] A. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations," *Phys. A Stat. Mech. its Appl.*, vol. 311, no. 3–4, pp. 590–614, 2002.

[136] M. E. J. Newman, "Coauthorship networks and patterns of scientific collaboration.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101 Suppl , pp. 5200–5205, 2004.

APPENDIX

A. Glossary of Terms

Below is a succinct list of terms related to Social Nework Analysis:

@

A @ is used to call out username names in tweet. When a username is preceded by the @ sign, it becomes a link to a Twitter profile.

 # (hashtag)

This is used to mark keywords or topics in a tweet.

Actor

A person, organization, or nation that is involved in a social relation.

Betweenness centrality

Is the measure of all geodesics between pairs of other vertices that include that vertex. Essentially, it is a measure of the number of times an actor connects pairs of other actors who otherwise would not be able to reach one another. It is a measure of the potential for control as well as the flow resources.

Blocking

To block someone means that they are unable to add you to their lists and as such you cannot view their updates or other information.

Blog

A frequently updated, chronologically ordered publication of personal thoughts and opinions with permanent links to other sources.

Clique

A set of connections in which each actor is directly connected to all other actors. In other words, a clique is a subnetwork with maximum density. Technically, cliques ideally contain a minimum of three actors.

Closeness centrality

This measures how close (based on path distance) on average an actor is to all other actors in the network.

Clustering coefficient

The clustering coefficient of a node is the ratio of existing links connecting the node's neighbors to each other, to the maximum possible number of such links. For nodes with fewer than two neighbors the clustering coefficient is undefined. The clustering coefficient of a node Z is 1 if every neighbor connected to Z is also connected to every other node within the neighborhood of Z, and 0 if no node that is connected to Z connects to any other node that is connected to Z. people.

Follower

Another Twitter user who subscribe to your tweets or updates.

Following / Friend

Twitter users you have chosen to subscribe to their tweets or updates.

Homophily

The principle that a contact between similar people occurs at a higher rate than among dissimilar people.

Mash-up

A web application that combines data from more than one source into a single integrated tool.

Podcasts / vlogs

Online audio and video logs that can be downloaded to devices such as PCs or handheld devices.

Power-law

A special kind a Mathematical relationship in which the distribution may have more sample data with extreme values than normal distributions. A power law distribution is characterized with a curve with a long tail lowering as the value increases.

Reciprocity

The action of returning similar acts.

Retweet

The act of forwarding another user's tweet to other to all of your followers.

RSS (Really Simple Syndication)

A family of web-feed formats used to push frequently updated content such as blog entries, news headlines, or podcasts to users' PCs or devices.

Social network

A social network is a social structure made up of individuals or organizations, commonly called actors or nodes, and the relationships between those actors, referred to as "ties."

Tweet

A message posted via Twitter containing 140 characters or less.

Widget

A small application that can be installed and executed within a web page by an end user.

Wiki

A publishing technology that is collaborative in nature for allowing multiple users to work on or publish documents online with appropriate version control.

CV

Jameson McFarlane

[redacted]

RESEARCH INTERESTS

Social Network Analysis; Information Security; Software Security Assurance; Big Data; Data Mining; Database Analysis; Software Engineering Methodology; Open Source Software

TEACHING INTERESTS

Open Source Software (PHP/MySQL/Apache, Linux); Programming; Database Management Systems; Systems Analysis and Design

EDUCATION

Doctor of Science in Applied Information Technology Candidate
Towson University, Towson, MD 2015

MBA in Computer Information Systems
York College of Pennsylvania, York, PA, 2006

B.Sc. in Computer Information Systems
York College of Pennsylvania, York, PA, 1997

AAS Teacher Ed. & Ed. Administration
Sir Arthur Lewis Com. College / University of the West Indies, 1990

TEACHING EXPERIENCE

Associate Professor
Computer and Networks Systems Administration, Thaddeus Stevens College of Technology 2005 – present
Courses taught:

- Database Management Systems; NET Programming (Windows and Web)
- Operation Systems (Linux Operating System);Management Information Systems
- Web Development in PHP/MySQL/ Apache; Exchange Server Administration

WORK EXPERIENCE

Applications Developer
Magnesita SA., May - August (2010, 2011)
Applications Developer

The Jay Group, Inc., summer 2008
Applications Developer

Mid-Atlantic Federal Credit Union, May – August 2007
Programmer Analyst

Computer Aid Inc., May – September 2005
Applications Developer / Systems Analyst

USA Direct Inc., May 03 - September 04
Programmer Analyst
State System of Higher Ed., Pennsylvania, 1998 - 2001

JOURNAL PUBLICATIONS

1. McFarlane, J. and Kaza, S. (2012), "State Governors on Social Media Reciprocity and Homophily in Twitter Networks," Conference for e-Democracy and Social Media, Singapore, Edition Danau-Universitat Krems
2. McFarlane, J. and Kaza, S. (2013), "Analysis of Research Collaboration Network at CISSE: Structure and Prominent Researchers," The 17th Colloquium for Information Systems Security Education (CISSE), Mobile, AL
3. McFarlane, J., Kaza, S., Bernard, L.(2015),(in submission),"The Social Identity of Cybersecurity Education Research: An Analysis of Publication and Collaboration from 1999 – 2013". IEEE Security and Privacy.

REFEREED CONFERENCE PUBLICATIONS
Extended Abstracts/Abstracts/Posters

1. McFarlane, J., Bernard, L., Kaza, S. (2015), "Social Identity of the Information Assurance Education Community: An Analysis of Publication and Collaboration from 1999-2013," The 19th Colloquium for Information Systems Security Education (CISSE), Las Vegas, NV, Poster
2. McFarlane, J. & Kaza, S. (2012), "Social Networking in Government: An Exploration of the use of the Twitter Micro-blogging service by U.S. State Governors," The 13th Annual International Conference on Digital Government Research (dg.o 2012), College Park, MD, ACM Press

HONORS AND ACHIEVEMENTS
Excellence in Teaching Award, Thaddeus Stevens College of Technology, 2011

TECHNICAL SKILLS
Programming Languages: Java, VB.NET, C#, R, PHP, Perl, COBOL, JavaScript, VBScript
Internet Technologies: AJAX, , HTML, XML, ASP.NET, jQuery
Databases: SQL/T-SQL, MS SQL Server, Oracle, MySQL, Sybase
Operating Systems and Environments: MS Windows, Linux