INTELLECTUAL STRUCTURE OF BUSINESS ANALYTICS

AND

DATA DRIVEN INSIGHTS FOR INFORMATION SECURITY BREACHES

by

RAHUL DWIVEDI

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2018

ACKNOWLEDGMENTS

"My journey is my destination" – a proverb I heard a decade ago, is true to my academic journey of the past four and a half years. I believe that this journey would not have been completed without the support and guidance of several individuals. I am wholeheartedly thankful to my dissertation chair, Dr. Sridhar Nerur, for his patience and guidance. I was greatly benefitted from discussions with him on various research problems. I am also thankful to Dr. Jingguo Wang, a member of my dissertation committee, for his insightful thoughts, and advice that had helped me to complete this work successfully. I am sure, the suggestions, guidance and constructive criticism I got had helped me to be a better researcher. Being from a non-statistics background, I am also thankful to Dr. Mary Whiteside, for providing me, with the necessary statistical knowledge from the courses I took with her. Statistics was never my cup of tea, but her courses along with masters and Ph.D. level courses I took with Dr. Mark Eakin, Dr. Mahmut Yasar and Dr. Douglas Grisaffe had made life easier for me.

For providing me the support in this journey, I would like to thank my wife, daughter, parents, in-laws and little son. Especially, my wife – Deepa, who took good care of our kids while I was working on my coursework and dissertation. My daughter – Udvita, who motivated me by emphasizing that once I am done with my Ph.D. and get a job, she will get more toys. The academic rigor and financial hardships during Ph.D. have taken a toll on my family, but eventually, I am able to succeed, and this success would not have been possible without my wife, daughter, and mother. It's hard to explain to my parents, what research I am doing, but they always wanted to finish me soon. My son, Divit, is too small to understand all these, but his cuteness has been a panacea in difficult times for me in the last few months.

I would also take this opportunity to thank the wonderful professors from whom I have learned so much during the Ph.D. research seminars. Dr. Radha Mahaptra, our Ph.D. program coordinator, has provided me with constant support and advice, especially during the rigorous journey of Ph.D. coursework. I am thankful to Dr. James Teng, Dr. Jennifer (Jie) Zhang and Dr. Riyaz Sikora, for introducing me to their respective fields of research and different research methodologies.

I would also like to thank my friends and colleagues, for their constant support. Kuldeep Singh, George Kurian, Bouabre Jean Baptiste Koffi, Jie Zhang, Sina Zare, Fereshteh Ghahramani and Dr. Adel Yazdanmehr – you guys will be missed. Last but not the least, I would like to thank Dr. Carolyn Davis, Peggy Schmitz, Evelyn Lucas, Nancy Morrel, Ashton Jones, Jennifer Hill, Samantha Juniker and Jennifer Lucas – some of whom are retired or not around – for helping me to carry out my GTA responsibilities, registering me for classes and helping me out with academic chores.

I hope that the proverb “my journey is my destination” remains true to the rest of my academic as well as non-academic life. I have been exposed to many research fields and have learned so much in the past four and a half years, thanks to all my professors and colleagues, which will keep me busy for at least the next few years. “Stay hungry, stay foolish” – another quote close to my heart by late Steve Jobs – is what I wish to aspire during the rest of my academic life. I hope that the creation of a void in my life because of the completion of Ph.D. will soon be filled with the research and teaching during my next academic endeavor.

ABSTRACT

Supervising Professor: Dr. Sridhar Nerur

Recent decades have exhibited phenomenal surges in not only in the amount of data generated globally but also in the methods used for analyzing data. Cybersecurity breaches have also increased in recent years. This dissertation explores these two important trends and topics in the information systems discipline: analytics and information security. In the first essay, I use a data science approach—analyzing research articles published in eight journals—to study the intellectual structure of business analytics (BA) within the information systems research community by analyzing research articles published in IS senior scholar's basket of eight journals. I employ citation count to identify reference disciplines, bibliographic coupling to clustering articles; and inter-citation counts to explore citation patterns. I also employ topic modeling to identify themes in the corpus of abstracts. Finally, I analyze social network using Exponential Random Graph Modeling (ERGM) to test the homophily effects for the co-authorship network. Information systems, computer science, general management, and economics are the most prominent reference disciplines. Predictive analytics, business intelligence, the Web, information technology (IT) management, firm performance, and decision support are important themes latent in the article abstracts. From the analyses of the co-authorship network for 184 unique authors, I found homophily based on continental affiliation (North American versus European institutions), departmental affiliations, and Ph.D.-granting institutions and university affiliation.

The second essay employs a data science approach to measure a firm's business relatedness and then tests its relationship with the firm's correlated risk in information security breaches. Using the Quadratic Assignment Procedure (MR-QAP) social network analysis (SNA) technique, I analyzed a network of 33 firms, all of which were breached within the last ten years. The data are available publicly in a dataset that documents cybersecurity breaches. Certain measures of firm similarities (business description and security risk factors) derived from the textual contents of their respective Securities and Exchange Commission (SEC) 10-K filings were found to be significantly correlated with their potential for breach. The similarity

of firms based on two-digit Standard Industry Classification (SIC) codes and research and development (R&D) expenditure (as a proxy for a firm's absorptive capacity) were also found to be significant.

The final essay performs specific and thorough analysis of insider security breaches, which entail security breaches carried out by current or past employees. This study aims to construct a theoretical understanding of insider breaches from the perspectives of employees. I draw on conservation of resources theory, social bonding theory, workplace deviance, and several motivation theories from organizational psychology to propose several hypotheses and to explain the results. I also consult reviews and ratings from a well-known job website, Glassdoor (www.glassdoor.com), for 71 public firms (40 breached and 31 non-breached). Analyses of textual reviews are carried out using IBM Watson's Tone Analyzer. Overall rating and rating for compensation and benefits from Glassdoor were found to be significantly correlated with the probability of breach using logistic regression analysis. My research also shows that the emotional tones of joy, fear, and anxiety are significant for an organization's potential to be affected by insider breach.

TABLE OF CONTENTS

List of Illustrations

List of Tables

Chapter 1

Introduction

"In God we Trust, all others must bring data."

~ Dr. William Edwards Deming

Dr. W. Edwards Deming, who provides the epigraph to this chapter, gave scholars the famous plan-do-check-act (PDCA) cycle for continuous improvement (Moen & Norman, 2006; Taylor et al., 2014). His aphorism, recorded a few decades ago, emphasizes the growing importance of data before the turn of the twenty-first century, and it is even more relevant today. With increasing amounts of data captured by organizations are working strenuously to form useful insights about consumers, competitors, and their own daily operations from data. The growing availability of data has prompted organizations to seek ways to gain competitive advantage by understanding these data and using them to make decisions (Davenport & Harris, 2007). "Business intelligence," "data science" and "data scientists," "big data," "data analytics," "business analytics," and "big data analytics" have become buzz phrases in the popular press[1]. Additionally, decreasing hardware costs and the prevalence of cloud services enable growth in data storage and usage by users. Increases in shared cloud space have enabled even small organizations and university classrooms, which formerly lacked the computational resources and storage capabilities to perform complex data-oriented tasks, to utilize big data to advantage. Massive Open Online Courses (MOOCs) offered by leading experts (mostly academics) have provided basic to advanced knowledge in the domain of data science to millions of individuals for no or minimal cost. In fact, as of March 2018, a search for the phrase "data science" on coursera.org, a well-known MOOC website, results in 1154 courses, including many specialized courses in data analytics. "Data scientist," "data engineer," and "data architect" are new high-

[1] https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html, https://www.nytimes.com/2017/08/25/dining/restaurant-software-analytics-data-mining.html, http://www.washingtonpost.com/sf/brand-connect/ibmpowersystems/wp/enterprise/big-data-turning-information-into-business-insights/, http://www.washingtonpost.com/sf/brand-connect/wp/2016/10/27/cit/analytics-puts-midmarket-companies-in-the-big-leagues/?utm_term=.73120dea5b84

paying position titles that have developed in response to increased capture, storage, and analysis of digital data. In 2012, the *Harvard Business Review* designated "data scientist" as the sexiest job of the twenty-first century.[2] Data analysis found its applications in domains as varied as human resources (HR analytics), healthcare (healthcare analytics), marketing (marketing analytics and consumer behavior analytics), and of course information systems and computer science (business and data analytics, web analytics, and social network analytics). The various applications of neural networks and deep learning[3] include self-driving vehicles (such as cars, drones and trucks), autonomous robots for various daily tasks, and smart devices, including intelligent personal assistants. These technologies are significant not only in the field of business but in research communities as well. Therefore, many experts consider an education in the tools and techniques of data analytics to be of utmost importance in terms of their applications to nearly any other field.

"Data science," "data analytics," and "business analytics" are popular phrases used almost synonymously by researchers from the information systems and computer science communities. According to (Dhar, 2013), data science is the study of "generalizable extraction of knowledge from data." More specifically, data analytics is the process of gaining useful and actionable insights based on a problem definition with the aid of statistical models as applied to existing data (Cooper, 2012). Additionally, BA is a movement or culture wherein either 1) fact-based decision-making is encouraged and rewarded; 2) a collection of meaningful techniques and practices gains useful insights from data; 3) data drives a transformational process aimed at better decision-making within an organization; or 4) a dataset leads to the development of organizational capabilities based on descriptive, predictive, and prescriptive models for improved decision-making (Holsapple et al., 2014). According to (Davenport & Harris, 2007), descriptive analytics describes or analyzes past events, opinions, or ideas (2007). Predictive analytics, on the other hand, attempts to

[2] https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

[3] The most notable application, which gained extensive media attention, is AlphaGo, a computer program not only capable of playing the board game, "Go," but also able to defeat a human professional. https://www.theatlantic.com/technology/archive/2016/03/the-invisible-opponent/475611/

determine future events based on the past. Finally, prescriptive analytics suggests certain actions based on available data. The research methods in this dissertation are primarily descriptive with some overlap in predictive analytics.

Apart from data analytics, this dissertation concentrates on cybersecurity. Specifically, I am interested in information security breaches. Although the two terms—cybersecurity and information security—are often used interchangeably, according to (von Solms & van Niekerk, 2013) the boundaries of cybersecurity encompasses much more than information security. As per International Telecommunications Union or ITU, cybersecurity is "the collection of tools, policies, security concepts, security safeguards, guidelines, risk management approaches, actions, training, best practices, assurance and technologies that can be used to protect the cyber environment and organization and user's assets."[4] On the other hand, information security protects and preserves the confidentiality, integrity, authenticity, reliability, and availability of information along with ensuring accountability of entities for their actions.[5] Thus, one can view information security as a subset of cybersecurity that protects information at the individual, computer, and computer network level. In contrast, cybersecurity itself operates in organizational and national spheres, which are the field's appropriate units of analysis. Hence, the research in this dissertation concerns cybersecurity rather than information security.

Cybersecurity attacks on government and private organizations result in heavy financial losses. For example, in a recent study by the Ponemon Institute, the annual global average cost of lost or stolen consumer records is $141 and this cost is $225 for United States-based organizations.[6] The same study also reported $3.62 million as the average total cost of data breach globally and $7.35 million for US-based firms. From an individual's perspective, increases in data breaches lead to growing incidents of identity theft. For example, a 2017 identity fraud study found that $16 billion was stolen from 15.4 million US-based consumers due to identity thefts. The same study also found that identity thieves within the US have

[4] https://www.itu.int/en/ITU-T/studygroups/com17/Pages/cybersecurity.aspx
[5] http://www.praxiom.com/iso-27000-definitions.htm#Information_security
[6] https://www.ibm.com/security/data-breach/#reports

stolen over $107 billion in the past six years.[7] To counter breach incidents, cybercrimes, and identity thefts, both government agencies and private organizations are constantly investing in cybersecurity research. A recent report by Gartner predicts global spending on information security to reach $90 billion in 2017 and to surpass $113 billion by 2020.[8] Similarly, Escal Institute of Advanced Technologies, popularly known as SANS institute's report on IT security spending trends found that budget for security spending as a percentage of IT budget is continuously increasing since fiscal year 2014.[9]

This dissertation explores the intersection of data analytics and cybersecurity – as cybersecurity analytics. Security analytics, in general, use the techniques of data science and analytics to gain useful information with the aim of preventing cyber-attacks (Mahmood & Afzal, 2013). The objective of this project is to advance scholars' understandings of the cybersecurity breaches—specifically, insider breaches—by using the techniques of popular data analytics. My research has three parts: first, I concentrate on the intellectual structure of business analytics (BA) based on articles published in leading information systems journals in the past two decades. The other two parts apply data analytics techniques to cybersecurity breach data.

.

[7] http://www.iii.org/fact-statistic/identity-theft-and-cybercrime
[8] http://www.gartner.com/newsroom/id/3638017
[9] https://www.sans.org/reading-room/whitepapers/analyst/security-spending-trends-36697

Chapter 2

Intellectual Structure of Business Analytics within the Information Systems Research Community: Evidence from the Senior Scholars' Basket of Eight Journals

2.1 Abstract

Business analytics and data science are consistently increasing topics of interest among academic researchers and industry practitioners of information systems. This chapter examines the theoretical underpinnings of studies of business analytics (BA) by using citation analyses, text analyses, and social network analysis to identify patterns in articles published in the eight leading journals in the field over the past 20 years (1997 to 2016). I conclude that in addition to information systems (IS), general business, organization science, and computer science are the major reference disciplines. The influence of IS on these studies is increasing recently in comparison with other disciplines. The influence of reference disciplines differs among the journals. North American journals and European journals also show distinct citation patterns. From text analyses, I extract primary research themes. Keyword frequencies change over time and differ across journals and journal groups. Finally, social network analysis of co-authorship networks found evidence of homophily based on certain author characteristics. The study provides an intellectual landscape based on the leading research in business analysis from the field of information systems.

2.2 Introduction

The phrases, "business analytics" and "data science," encompass a variety of business intelligence tools and applications. For example, website analytics involves analyzing the browsing behavior of online users to gain insights into their past behaviors and also to predict their future actions. Similarly, supply chain analytics involves applying analytics capabilities to gain operational efficiencies at the firm level, and HR analytics involves finding the right match between job roles and personal capabilities of prospective job candidates. Businesses and researchers are becoming more interested in analytics partly because large-scale data and the tools for analyzing these data—including cloud computing and cloud-based services—are becoming more available and more affordable. For example, the technology website Gizmodo[10] shows that 300 million photos—or, 500 terabytes of data—are uploaded on Facebook every day. Similarly, on average, Google processes over 40,000 search queries every second.[11]

Researchers analyze these data rigorously for diverse purposes, which range from gaining useful customer insights, providing product recommendations to customers, making better business decisions (sometimes in real time), and devising better marketing strategies to target end customers. For example, a survey conducted by BloomReach in 2015 on 2000 regular online shoppers in the US found that 44% of customers visit amazon.com directly to find products instead of searching through engines like Google, Yahoo, or Bing.[12] Survey participants cited that their main reason for choosing Amazon over search engines is Amazon's unique capability to provide personalized product recommendations based on advanced personalization algorithms.

A similar study published in the *Harvard Business Review*, which was based on a survey conducted with Chief Information, Analytics, Marketing, and Data Officers representing *Fortune* 1000 US firms, reported

[10] https://gizmodo.com/5937143/what-facebook-deals-with-everyday-27-billion-likes-300-million-photos-uploaded-and-500-terabytes-of-data
[11] http://www.internetlivestats.com/google-search-statistics/
[12] https://www.bloomreach.com/en/resources/blogs/2015/10/amazon-commands-nearly-half-of-consumers-first-product-search.html

that about 80% of executives characterized their investments in big data as successful.[13] Decreasing expenses and increasing revenues, finding novel innovation avenues and launching new products/services, transforming business for the future, increasing the speed of current efforts, and establishing a data-driven culture are the major areas where executives reported to see the benefits of their big data investments.

This growing interest in data analytics by industries paralleled a proliferation of academic articles published in the same field. A quick search in the Web of Science academic database for the phrases, "business analytics," "business intelligence," "data analytics," "data science," "predictive analytics," "machine learning," "deep learning," and "analytics" produced 54819 results—those are, articles and conference proceedings published in English between 1990 and 2016, with the majority of articles published in the past few years.

As a discipline, information systems (IS) bridges computer science, business management, and social sciences. IS scholars concentrate on the use and usefulness of technological capabilities of systems of people, ideas, and information in organizational work. Historically, IS involved modeling data, designing databases, understanding communications and information processing technologies, and knowing how applications and services affect individuals and businesses (Davis, 2006). That background, along with the data surge regarding both consumers and businesses, places IS at the crux of analytics research. The field of IS negotiates between the techniques of computer science and those of business problem-solving. Thus, IS researchers play important roles in advancing the body of knowledge of BA research.

One of the purposes of this chapter is to examine the theories of BA research conducted by IS researchers by analyzing the articles in eight industry-leading journals published from 1997 to 2016. My methodology includes citation analyses and text analyses to answer the following questions: 1) What are the primary reference disciplines for business analytics? 2) With what frequency are the identified reference disciplines cited over time? 3) Is inter-citation homophily evident per reference discipline and/or per journal group

[13] http://newvantage.com/wp-content/uploads/2017/01/Big-Data-Executive-Survey-2017-Executive-Summary.pdf

(i.e., North American versus European journals)? 4) What thematic patterns are apparent in the corpus? 5) Are thematic patterns constant or evolving over time? and 6) Are specific journals invested in specific themes within IS research, or do industry-wide themes appear across all journals?

After performing citation counts for the BA articles published in leading IS journals, I identified eleven reference disciplines. The top reference disciplines include general business, organization science, and computer science. The influence of IS is increasing over the past few years compared with other reference disciplines. Economics as a reference discipline has no influence on Information Systems Journals (ISJ) and European Journal of Information Systems (EJIS) while Management Information Systems Quarterly (MISQ), Journal of Association of Information Systems (JAIS), Information Systems Research (ISR), and Journal of Management Information Systems (JMIS) cite more computer science journals. In addition, European journals cite more IS journals than North American journals. When comparing them to groups for inter-citation counts, I found that European journals cite North American journals more often than vice versa.

Testing for the homophily principle shows that affiliations based on authors' continental affiliations, departmental affiliations, affiliations with Ph.D.- granting institutions and universities are significant predictors of co-authorship links. Authors' genders and post-Ph.D. experience were insignificant. The insignificance of gender suggests collaborations between men and women scholars publishing BA articles in leading IS journals. Also, the insignificance of experience as a predictor of scholarly collaboration suggests that recent postdoctoral researchers collaborate with colleagues with long research histories in their fields.

Based on the text analysis technique of topic modeling, I learned that predictive analytics, business intelligence, the Web, IT management, firm performance, and decision support are primary themes in BA IS research. In addition, the keyword frequency method shows that "decision support systems," "knowledge," and "web" were prominent keywords before 2010. Between 2011 and 2013, "data" and "financial models" are used more frequently. Usage of the terms, "big," "data," and "analytics" became frequent in BA IS publications after 2013. Also, the increased popularity of keywords, “data” and “business” suggest scholars’ growing interests in organizations in IS studies. Across journal groups, “web,” “predictive,” “model,” and “markets” were terms used often in North American journals whereas Business Process Outsourcing (BPO),” “big,” Enterprise Resource Planning (ERPII),” and “management” were mostly used in European journals.

2.3 Literature Review

This study uses three main research methodologies for investigating and defining the intellectual structure of business analytics within the information systems domain: citation analysis, text analysis, and social network analysis. My research supplements extant studies, noted below, that have helped scholars explore business analytics from the IS perspective.

2.3.1 Citation Analysis

Citation analysis demonstrates a broad perspective on scholarly communication using large datasets (Borgman, 1990). Researchers have historically used the methodology to evaluate scholarly contributions of to various disciplines and to determine knowledge flows, the diffusion of ideas, and the intellectual structure of scientific disciplines (D. Zhao & Strotmann, 2015). Citation studies are based on the assumption that studies constitute a research theme if they cite each other; if they are frequently cited together; or if they have many cited references in common (D. Zhao & Strotmann, 2015). Thus in citation studies, relationships between keyword clusters are measured using methods, such as inter-citation counts (Boyack, Klavans, & Börner, 2005), co-citation counts (H. D. White & Griffith, 1981, 1982; H. D. White & McCain, 1998), and bibliographic coupling counts (D. Zhao & Strotmann, 2008). Inter-citation counts determine the number of times two objects have cited each other; co-citation counts calculate the frequency of documents citing two objects together; and bibliographic coupling refers to the number of cited references that two objects have in common (D. Zhao & Strotmann, 2015).

The intellectual structure of a research field includes the characteristics, patterns, organizations, and temporal evolution of the discipline, so its scholarly communities frequently apply citation analyses techniques (Culnan, 1987; Eom, 1996; S. P. Nerur, Rasheed, & Natarajan, 2008; Ramos-Rodríguez & Ruíz-Navarro, 2004). For example, citation or bibliometric analysis techniques have been used to study the intellectual structure of strategic management (S. P. Nerur et al., 2008; S. Nerur, Rasheed, & Pandey, 2015), information science (H. D. White & McCain, 1998), economics (McCain, 1991), management information system (Culnan, 1986, 1987), supply chain management (Charvet, Cooper, & Gardner, 2008), organizational behavior (Culnan, O'Reilly, & Chatman, 1990), operations management (Pilkington & Meredith, 2009), decision support systems (Eom, 1996), and information security (Olijnyk, 2015).

(Boyack & Klavans, 2010) provide a detailed comparison of major citation analyses techniques for a large corpus of biomedicine literature. Similar comparison studies were carried out by (Jarneving, 2005) and (Shibata, Kajikawa, Takeda, & Matsushima, 2009).

The first step in citation analysis collects a set of citing articles or authors whose research exists within a specified time-period (McCain, 1990; D. Zhao & Strotmann, 2015). This dissertation collects business analytics articles published between 1992 and 2016 in eight peer-reviewed journals, which are indexed in ISI's Web of Science citation database. To analyze these articles from a citation analysis perspective, I use bibliographic coupling, which was first proposed by (Kessler, 1963), who defined a unit of coupling between two research articles as the common item in the respective reference list of the two. Subsequent papers constitute the same group if each member of the group has at least one coupling unit to another paper belonging to the group. Thus, two documents are bibliographically coupled if they contain the same item in their list of references. The coupling strength between articles is measured by the number of coupling units between them. That is, the number of items the two shares in their reference list is the bibliographic coupling frequency for the two. Higher frequency numbers indicate greater relatedness between articles (D. Zhao & Strotmann, 2015). Finally, several articles constitute a related group if each member of the group has at least one coupling unit to every other member of the group (Kessler, 1963).

(D. Zhao & Strotmann, 2008) find several advantages of bibliographic coupling as a citation technique in mapping the intellectual structure of a scientific discipline. The foremost advantage is the methodology's capacity to enable scholars to place temporal boundaries on research objects. In contrast with co-citation analysis, which requires researchers to consider an entire citation bank, bibliographic coupling is more appropriate to fields like business analytics, which evolve rapidly.

Analyses performed for this chapter used bibliographic coupling to find similar studies and groups them based on the commonly-cited references. These document clusters expose different research themes studied by IS researchers focusing on BA. Further, I also compare the citation patterns for the leading North American versus European information systems journals.

Research conclusions suggest that that citation analysis techniques alone provide few insights into defining the characteristics of a discipline and should be considered with other techniques (Balijepally & Nerur, 2015). Therefore, this chapter concurrently employs text analysis, topic modeling, and social network analysis of co-authorship networks to learn more about the intellectual structure of BA within IS. Also, considering authors as units of analysis for citation studies opens the possibility of exploring questions related to the social structure of a discipline in addition to its intellectual structure (H. D. White, 1990).

2.3.2 Text Analysis

Text mining, also known as text data mining or knowledge discovery, refers to the process of extracting interesting and non-trivial patterns or knowledge from text documents (Tan & others, 1999). Text mining is considered a sub-field of data mining. Whereas data mining finds useful information patterns in databases containing structured, semi-structured, or unstructured stored data (Fan, Wallace, Rich, & Zhang, 2006), text mining uses techniques from natural language processing to supplement knowledge discovery in databases, data mining, machine learning, and statistics (Hotho, Nürnberger, & Paa\s s, 2005). Scholars have also surveyed the benefits of various text mining techniques and methods such as clustering, classification, information retrieval, and extraction (M. W. Berry & Castellanos, 2004; Gupta, Lehal, & others, 2009; Hotho et al., 2005). Applications of these techniques to the World Wide Web resulted in the emergence of web mining to automatically discover and extract information from web documents and services. (Kosala and Blockeel, 2000) provide a brief survey of web mining research.

Recently, scholars of business, including fields such as marketing and consumer research, have utilized text analysis to understand the intellectual structure of their respective disciplines. For example, Mela, Roos, &

Deng (2013) analyze keywords to develop insights on the evolution of the journal, *Marketing Science*. Similarly, researchers carried out an historical analysis of 40 years of the *Journal of Consumer Research* using text mining to uncover key phrases and citation patterns (X. (Shane) Wang, Bendle, Mai, & Cotte, 2015).

2.3.3 Social Network Analysis

A social network consists of a finite set of actors and the relations defined among them (Wasserman & Faust, 1994). Social network analysis is a broad strategy for studying social networks including bibliometric networks (such as citation, co-citation, and co-authorship networks), patent citation networks, transport networks, and other forms of interaction networks among individuals, organizations, or nations (Otte & Rousseau, 2002). Freeman (2004) lists four essential features for social network analysis: 1) motivation from a structural intuition based on social ties; 2) grounding in systematic empirical data; 3) reliance on graphic imagery; and 4) use of mathematical and/or computational models. Although historically developed as a sub-field of sociology, social network analysis now includes techniques from mathematics and computer science, including graph theory. Numerous studies provide details on the historical developments, terminology, and techniques used for performing social network analysis (S. P. Borgatti, Mehra, Brass, & Labianca, 2009; Freeman, 2004; Scott, 2012; Wasserman & Faust, 1994). UCINET for Windows platform, R for both Windows and Linux environments, and Pajek are currently the most popular software for carrying out social network analysis (Stephen P Borgatti, Everett, & Freeman, 2002; De Nooy, Mrvar, & Batagelj, 2011).

Social and information scientists' interests in this field are growing due to the data implications of popular social networking websites such as Facebook and Twitter, which increase digital interactions between users. Hence, scholars from different fields such as sociology, computer science, psychology, information systems, and mathematics examine these websites hoping to understand users' practices, implications, and cultures (Boyd & Ellison, 2007).

2.3.4 Homophily

Social correlation is the correspondence between the behavior of affiliated members in a social network—i.e., for two nodes *u* and *v* in a graph G, the event that *u* becomes active is correlated with *v* becoming active (Anagnostopoulos, Kumar, & Mahdian, 2008). This correlation between members of a social network can be explained by the phenomenon of homophily. Homophily is the principle that contact between similar people occurs at a higher rate than among dissimilar people (McPherson, Smith-Lovin, & Cook, 2001). Homophily as a basic organizing principle for network formation has been studied since the 1920s and 1930s by educational and development psychologists (Freeman, 1996). ((McPherson et al., 2001) provide a detailed historical account of the different types of relationships studied through the lens of homophily to explain individual behavior, attitudes, abilities, beliefs, and aspirations. Those relationships include such based on race, ethnicity, gender, age, religion, education, occupation, and social class. ((McPherson et al., 2001) found that geography, family ties, organizational affiliations, and cognitive processing also cause homophily.

2.4 Data Collection

Research articles comprise the raw data in this chapter, and I collected them from the Web of Science Core Collection Indexes. My objective is to identify the business analytics articles published in the IS senior scholars' basket of eight journals[14] between 1997 and 2016. I used the following search keywords: "business analytics," "business intelligence," "data Analytics," "data Science," "predictive analytics," "machine learning," "deep learning," and "analytics," which resulted in 68 articles. Figure 2.1 shows article distribution over the years indicated as published in the IS senior scholars' basket of eight journals.

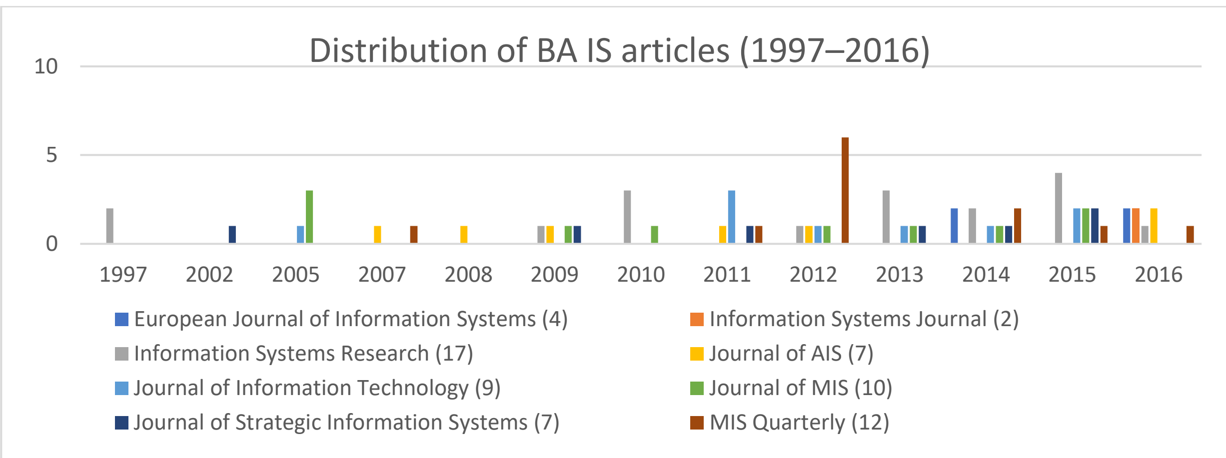


Figure 2.1: Distribution of BA IS articles by IS senior scholars' basket of eight journals over the last 20 years

[14] https://aisnet.org/page/SeniorScholarBasket

2.5 Research Methodology

2.5.1 Exponential Random Graph Modeling (ERGM)

One of the main assumptions in social science research based on random sampling of individuals is that those sampled do not interact with one another. This generalization is popularly known as the assumption of independence in regression parlance. Traditional regression models based on this assumption cannot be applied to test for statistical relationships in the context of social networks where nodes (individuals or organizations) interact with one another often (Harris, 2013a). Exponential random graph models (ERGMs) are a class of tie-based statistical models for social networks that account for the presence or absence of network ties and so provide a model for network structure (Lusher, Koskinen, & Robins, 2012). Therefore, ERGM is a tool for examining relationship patterns and for identifying the ways in which network members' characteristics and broad cultural influences can explain or predict observed relationship patterns (Harris, 2013a). For analysis of co-authorship networks, I chose to use "statnet" and "ergm" packages of R statistical language because these packages represent, visualize, analyze, appropriate, and simulate the co-authorship network under analysis (Handcock, Hunter, Butts, Goodreau, & Morris, 2008; Hunter, Handcock, Butts, Goodreau, & Morris, 2008a).

2.6 Analysis and Results

2.6.1 Identification of reference disciplines for Business Analytics research published in leading Information Systems journals

Citations from the 68 articles were analyzed to identify the reference disciplines for the BA IS research articles published in selected journals. Business analytics articles produced 5031 citations from which 47 cannot be identified as belonging to any known cited reference. As a result, I found 4984 cited references in 2036 distinct publications with a mean of 73.3 cited references per article. Because most references are cited just once, my data are skewed. Therefore, I we performed the following data reduction procedure before identifying reference disciplines: I removed all journals cited less than ten times between 1997 and 2016. For the 68 articles, this step resulted in 2151 citations, which I used to identify common reference disciplines. In classifying the list of cited references into reference disciplines, I was consistent with prior literature wherever possible (Agarwal, 2016; Grover, Gokhale, Lim, Coffey, & Ayyagari, 2006). I also consulted a master list of journals from Web of Science, which identifies 11,144 indexed journals into 22 disciplinary categories.

The 11 referenced disciplines with important journal outlets are shown in Table 2.1 with the distribution in Figure 2.2:

Discipline	Journal name
Computer Science	Communications of ACM, IEEE Transactions on Knowledge Data Engineering, Machine Learning, IEEE Intelligent Systems, Artificial Intelligence, Journal of Machine Learning Research, Lecture Notes in Computer Science, ACM Transactions on Information Systems.
Economics	American Economic Review, Quarterly Journal of Economics
General Business	Management Science, Harvard Business Review, Decision Science, Sloan Management Review, MIT Sloan Management Review, California Management Review
General Science	Science, Proceedings of the National Academy of Sciences
Information Systems	MIS Quarterly, Information Systems Research, Journal of Management Information Systems, Decision Support Systems, Journal of Strategic Information Systems, Journal of Information Technology, European Journal of Information Systems, Journal of Management Information Systems, Journal of Association of Information Systems, Information Management, Information Systems ,Information Systems Journal, Information & Management, Communications of Association of Information Systems, Information Systems Management, Journal of American Society of Information Science and Technology,, Journal of American Society of Information Science, Business Intelligence, MIS Quarterly Executive, Expert Systems With Applications, International Journal of Information Management, Information Processing & Management
Marketing	Marketing Science, Journal of Marketing Research, Journal of Marketing, Journal of Academy of Marketing Science, Journal of Consumer Research
Operations Research / Operations Management	European Journal of Operations Research, Journal of Operations Management
Organization Science	Organization Science, Academic of Management Review, Strategic Management Journal, Academy of Management Journal, Administrative Science Quarterly, Journal of Management
Psychology	Journal of Personality and Social Psychology, Group Decision & Negotiation, Psychological Review
Sociology	American Journal of Sociology, American Sociological Review
Statistics	Statistical Science
Working papers	Others

Table 2.1. Reference disciplines for BA IS publications for IS senior scholars' basket of eight journals

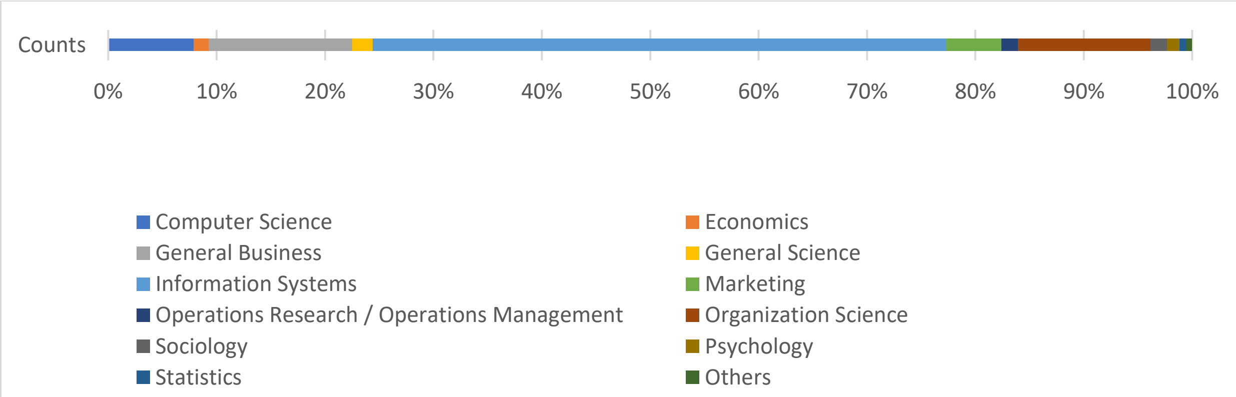


Figure 2.2. Proportions of referenced disciplines for BA IS articles published in IS senior scholars' basket of eight journals

As table 2.1 and figure 2.2 show, most articles cite information systems journals (1137 cited references) as reference journals followed by general business (284), organization science (262), and computer science (162). As noted above, the total number of cited references is 2151.

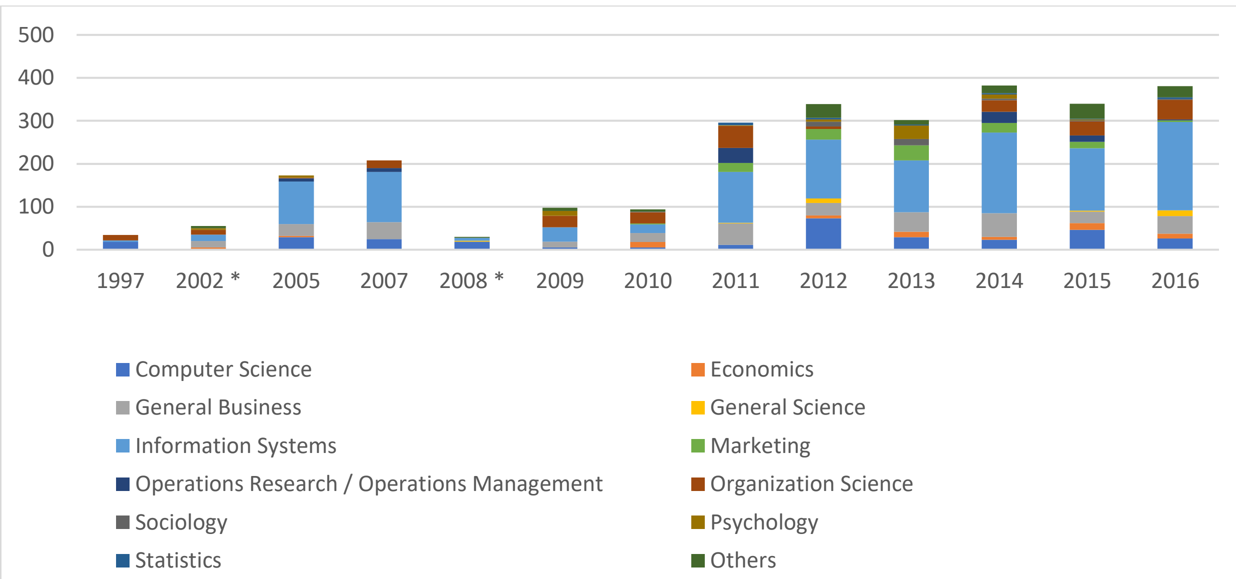


Figure 2.3. Proportional distribution of cited reference disciplines, 1997–2016

Figure 2.3 shows that information systems is the most cited reference discipline annually aside from years 2002 and 2008, when only one publication was avalable in the sample. Out of 3036 total cited references, 1273 belongs to information systems journals, followed by 429 for general business, 404 for computer science and 278 for organization science journals. Again, I did not include the journals that were cited only once. Most cited reference disciplines include, in order, information systems followed by computer science, organization science, and general business.

2.6.2 Analysis of inter-citation counts

Inter-citation counts provide evidence as to whether cited reference disciplines differ in selected journals. Figure 2.4 shows the distribution of references disciplines across journals.

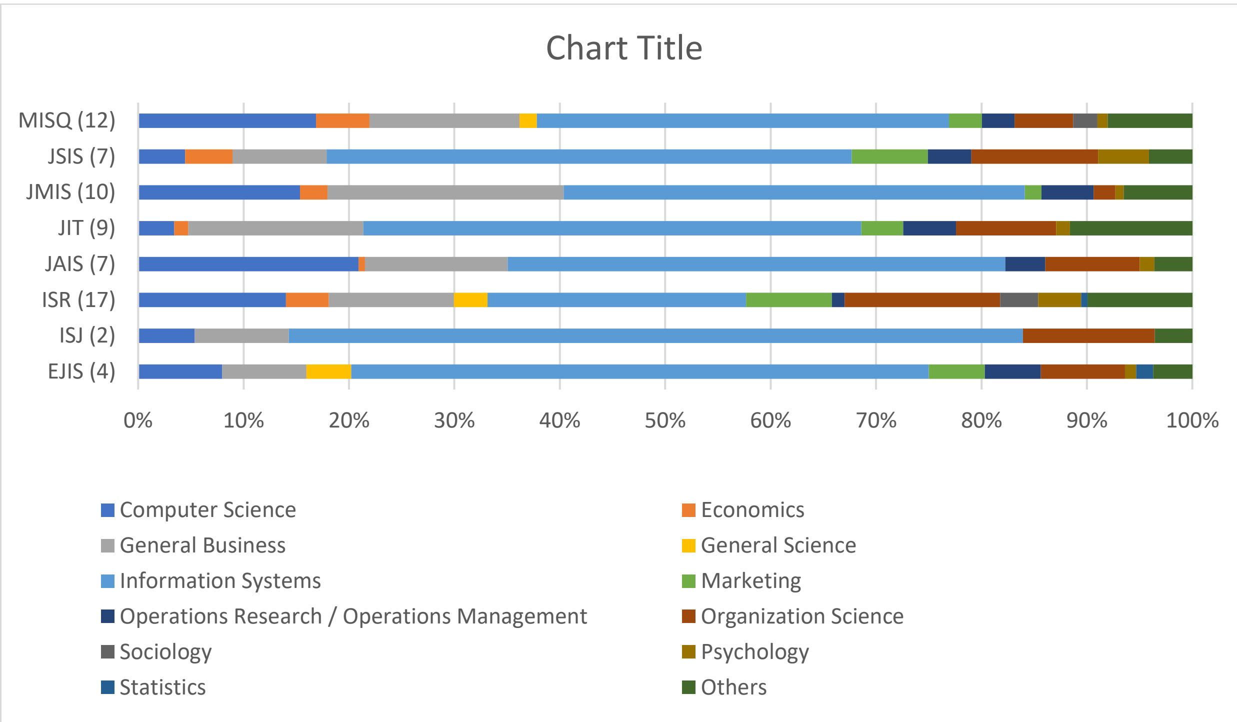


Figure 2.4. Distribution of reference disciplines across journals

Chi-square test of independence is carried out to determine whether the likelihood of citation counts for reference disciplines across journals is random or significant. I hypothesized that the underlying patterns of citation counts for the reference disciplines is significant and that the actual values differ significantly from the expected values. Thus, journal categories and respective citation counts for reference disciplines would not be independent. The p-value derived from performing the chi-square test ($p<<0.0001$) proves that the distribution of citation counts for reference disciplines across journals are not random and therefore are significantly statistically different from expected counts.

Also noteworthy are differences in citation frequencies for top North American IS journals and their European counterparts. From our data, we see that from 3036 cited references, 29.84% (906 cited references) come from European IS journals and 70.16 % (2130 cited references) from North American IS journals. Table 2.2 and Figure 2.5 show the percentage distribution of cited reference disciplines for the European and North American IS journals.

	CS	Econ	GB	IS	Mark	OR / OM	OS	Sociology	Psychology	Statistics	Others
European IS Journals (EJIS, JIT, JSIS and ISJ)	4.86	1.99	12	51.435	5.08	4.525386	10.26	0	2.3178808	0.331126	7.1744
North American IS journals (MISQ, ISR, JMIS and JAIS)	16.9	3.33	15	37.887	3.71	3.051643	8.685	1.784038	2.0187793	0.187793	7.4178

Table 2.2 Percentage distribution for reference disciplines for 906 cited references in European IS journals and 2130 cited references in North American IS journals

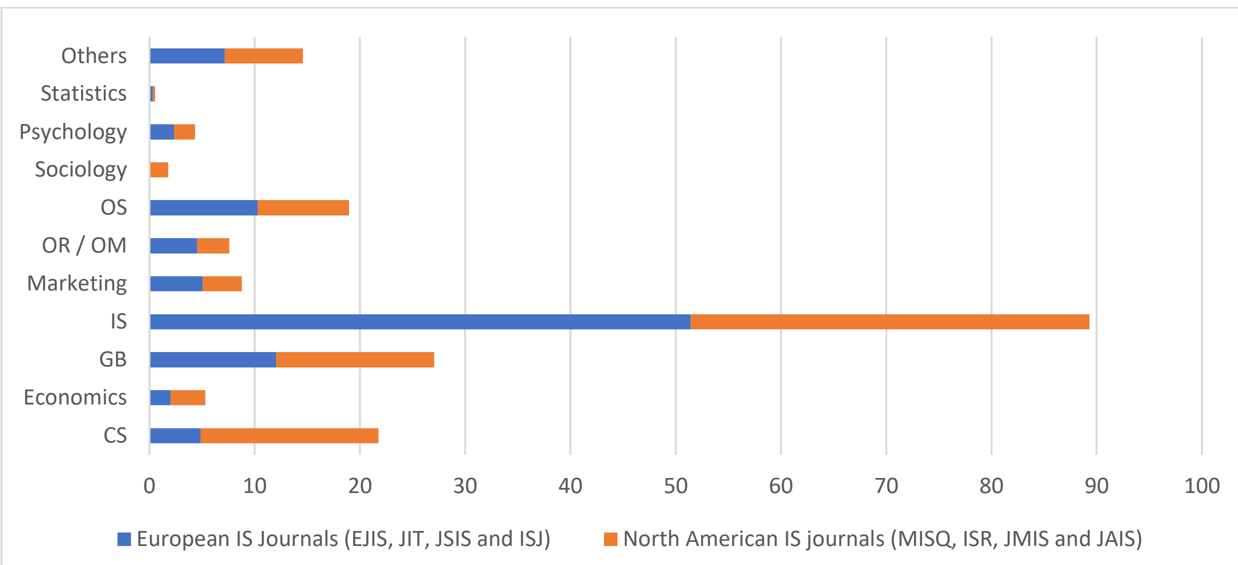


Figure 2.5 Percentage distribution for reference disciplines for 3036 cited references

For brevity, Table 2.2 and Figure 2.5 employ the following acronyms: Organization Science (OS), Operations Research/Operations Management (OR/OM), Information Systems (IS), General Business (GB), and Computer Science (CS). Those figures show little difference in reference disciplines for leading North American versus European IS journals regarding business analytics research. This conclusion is further supoprted by a t-test carried out on the percentage of cited references for the two groups, which found the p-value to be greater than 0.1 (i.e. p >> 0.1).

My final conclusion regarding reference disciplines maps IS citation patterns among top IS journals themselves. From a total of 1236 IS references, 805 belongs to IS senior scholars' basket of eight journals citing each other. As expected, *MIS Quarterly* is most cited IS journal with 351 citations, followed by *Information Systems Research* with 141 citations and *JMIS* with 132 citations. *Information Systems Journal* is the least cited journal with only 15 citations. The distribution in Figure 2.6 summarizes these results.

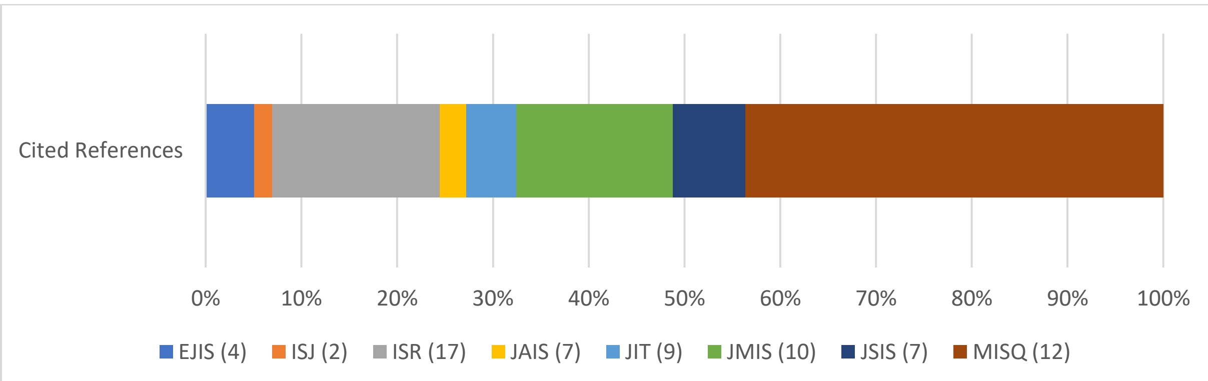


Figure 2.6 Percentage distribution of cited references for IS senior scholars' basket of eight journals

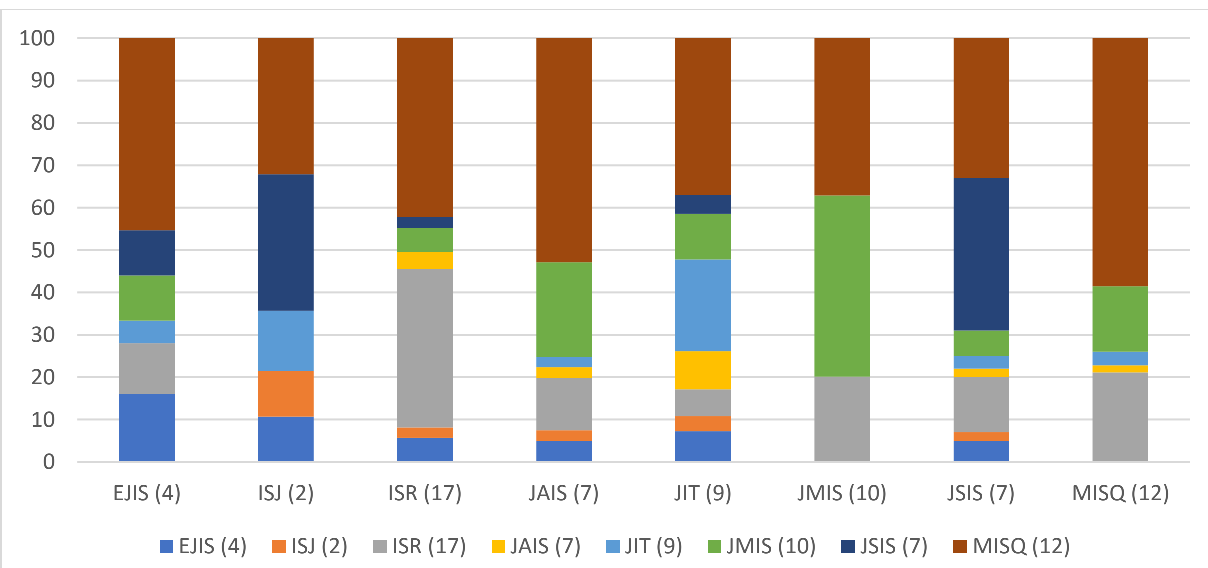


Figure 2.7 Inter-citation proportional distribution of basket of eight BA IS journals

Figure 2.7 and Table 2.3 show the corresponding percentage distribution for cited references for IS journals citing each other.

	EJIS (4)	ISJ (2)	ISR (17)	JAIS (7)	JIT (9)	JMIS (10)	JSIS (7)	MISQ (12)
EJIS (4)	16	0	12	0	5.33	10.6667	10.667	45.33333
ISJ (2)	10.71	10.71	0	0	14.3	0	32.143	32.14286
ISR (17)	5.691	2.439	37.398	4.065	0	5.69106	2.439	42.27642
JAIS (7)	4.959	2.479	12.397	2.4793	2.48	22.314	0	52.89256
JIT (9)	7.207	3.604	6.3063	9.009	21.6	10.8108	4.5045	36.93694
JMIS (10)	0	0	20.161	0	0	42.7419	0	37.09677
JSIS (7)	5	2	13	2	3	6	36	33
MISQ (12)	0	0	21.138	1.626	3.25	15.4472	0	58.53659

Table 2.3 Percentage distribution for IS senior scholars citing each other

In table 2.3, the rows contain the number of cited references for particular IS journals, which cross-reference columns correspond to other journals. For example, the first row designates cited references in *EJIS*, of which 16% are for other *EJIS* articles, 12% for *ISR* articles, 5.33% for *JIT* articles, 10.67% for *JMIS* articles, 10.67% for *JSIS* articles, and 45.33% for articles published in *MIS Quarterly*. Concurrently, columns represent journal citations vertically. Therefore, diagonal of Table 3 represents self-citations: 58.53% for *MIS Quarterly*, 37.398% for *ISR*, and 36% for *JSIS*. See, importantly, that *JMIS* cites only itself, *ISR*, and *MIS Quarterly*. Table 3 shows that *MISQ* is the most cited IS journal among for BA IS research among all. It is also the journal with the highest percentage of self-citations.

Finally, I separated citation analyses for top IS journals by North American versus European continental affiliation. Figure 2.8 demonstrates that approximately 40% of cited references in business analytics articles in top European IS journals reference themselves while 60% cite North American journals. Conversely, approximately only 6% of cited references in North American journals are dervied from articles published in European journals, so 94% of North American IS articles reference only other North American journals. The significant p-value for the chi-square test ($p<<0.01$) reinforces this conclusion.

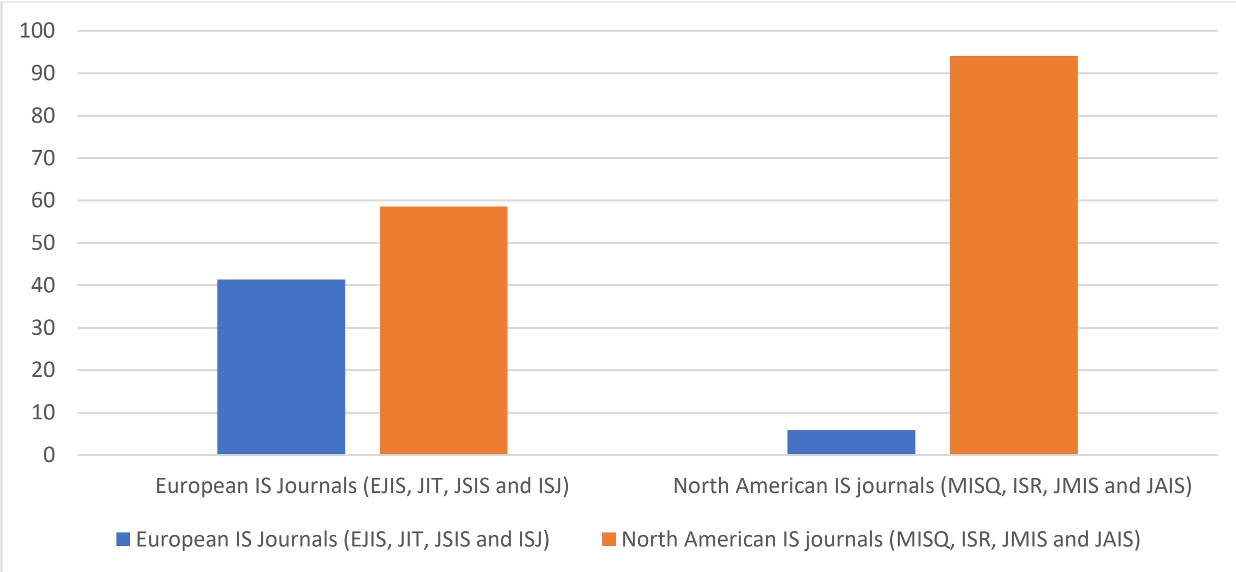


Figure 2.8. European versus North American journals: How top IS journals for BA research cites each other

2.6.3 Clustering articles using bibliographic coupling

As mentioned above, bibliographic coupling suggests the similarity of documents based on overlapping references cited by those documents. That is, if articles A and B cite the same article C, then A and B share disciplinary topics. Bibliographic coupling analysis of journals analyzed in this dissertation supplies the themes shown in Table 2.4.

Cluster No.	Articles	Underlying Themes
Cluster 1:	(Abbasi, Albrecht, Vance, & Hansen, 2012), (Bardhan, Oh, Zheng, & Kirksey, 2015), (Chellappa, Sambamurthy, & Saraf, 2010), (R. Clarke, 2016), (Fayard, Gkeredakis, & Levina, 2016), (Ghoshal, Menon, & Sarkar, 2015), (X.-B. Li & Sarkar, 2014), (Loebbecke & Picot, 2015), (Newell & Marabelli, 2015), (Pant & Sheng, 2015), (Raghunathan & Sarkar, 2016), (Shmueli & Koppius, 2011), (Zheng, Fader, & Padmanabhan, 2012)	Predictive analytics, Web, Crowd, Markets, Decision making.
Cluster 2:	(Baker, Jones, & Burkman, 2009), (Choudhary & Vithayathil, 2013) (Dong, Huang, Sinha, & Xu, 2014), (Greenwald, Kannan, & Krishnan, 2010), (Kayande, De Bruyn, Lilien, Rangaswamy, & van Bruggen, 2009), (Meyer et al., 2014), (Pant & Srinivasan, 2010), (Pant & Srinivasan, 2013), (Zhu, Prietula, & Hsu, 1997)	Predictive analytics, Decision support, Business processes, Data.
Cluster 3:	(Audzeyeva & Hudson, 2016), (H. Chen, Chiang, & Storey, 2012), (Chung, Chen, & Nunamaker Jr, 2005), (Deng & Chi, 2012), (Fan, Gordon, Pathak, & PATHAK, 2005), (D. Hu, Zhao, Hua, & Wong, 2012), (Shollo & Galliers, 2016)	Business intelligence, Web, IS Use.
Cluster 4:	(Abbasi et al., 2015), (Jabr, Mookerjee, Tan, & Mookerjee, 2013), (Lacity, Solomon, Yan, & Willcocks, 2011), (Xin Li, Chen, Zhang, Li, & Nunamaker, 2009), (Petrini & Pozzebon, 2009), (Sharma, Mithas, & Kankanhalli, 2014), (Susarla, Barua, & Whinston, 2010)	Business process outsourcing, Business, and predictive analytics.
Cluster 5:	(D. Q. Chen, Preston, & Swink, 2015), (Coltman, Devinney, & Midgley, 2011), (Habjan, Andriopoulos, & Gotsi, 2014), (G. Kim, Shin, Kim, & Lee, 2011), (Xixi Li, Hsieh, & Rai, 2013), (Popovič, Hackney, Coelho, & Jaklič, 2014), (Wakefield, 2013)	Firm performance, IS Usage, SCM.
Cluster 6:	(Chau & Xu, 2012), (Constantiou & Kallinikos, 2015), (Gholami, Watson, Molla, Hasan, & Bjørn-Andersen, 2016), (Koh, Gunasekaran, & Goodman, 2011), (Lau, Liao, Wong, & Chiu, 2012), (Roussinov & Chau, 2008)	ERP, supply chain, Big data, Business intelligence.
Cluster 7:	(Abbasi, Sarker, & Chiang, 2016), (Agarwal & Dhar, 2014), (Müller, Junglas, vom Brocke, & Debortoli, 2016), (Park, Huh, Oh, & Han, 2012), (Provost, Martens, & Murray, 2015), (Sundararajan, Provost, Oestreicher-Singer, & Aral, 2013)	Information, Big data analytics, Privacy.
Cluster 8:	(Luftman & Zadeh, 2011), (Luftman et al., 2012), (Luftman et al., 2013), (Luftman et al., 2015)	IT management.
Cluster 9:	(Arnott & Pervan, 2005), (Arnott & Pervan, 2012), (Arnott & Pervan, 2014), (Rouibah & Ould-Ali, 2002)	Decision support systems.
Cluster 10:	(T. D. Clark, Jones, & Armstrong, 2007), (Mookerjee & Mannino, 1997), (Nelson, Todd, & Wixom, 2005)	Information, management support systems.
Cluster 11:	(H. M. Kim, Fox, & Sengupta, 2007)	Data models, Compliance.

Table 2.4. Clustering of documents based on Bibliographic coupling and underlying themes

In Table 2.4, Clusters 1, 2, 3 and 4 represent overlapping topics including web, predictive analytics, and business intelligence. As is evident, bibliographic coupling produces interrelated clusters without strict boundaries demarcated between topics.

Since the technique of bibliographic coupling does not provide researchers with neat clusters, and since the number of articles in my data set is limited, I instead cluster topics using the data mining technique known as K-means clustering.

2.6.4 Clustering articles using K-means algorithm

K-means clustering is a simple, yet powerful unsupervised data mining technique for classifying a set of *n* data points into K clusters. Here, the data points within a cluster are closer in distance to the centroid of that cluster compared with their distance from the centroids of other clusters.

To determine the optimum number of Ks, I followed the procedure outlined by Mueller and Massaron (2015) and used the K-Means clustering algorithm in "sklearn," a machine learning module in Python. The K-Means algorithm returns a measure called "inertia," which is the aggregate of the difference between every data point and the centroid of the cluster to which it belongs. A smaller value of inertia suggests a more cohesive cluster. Based on the rate of change of inertia against different values of K, I followed (Mueller & Massaron, 2015) by choosing the value of K that corresponded to the one that caused the biggest "jump" in rate. Thus, I chose seven clusters for analysis in this section.

Cluster No.	Articles	Underlying Theme
Cluster 1:	(Chung et al., 2005) (Pant & Srinivasan, 2010) (Chellappa et al., 2010) (Shmueli & Koppius, 2011) (Zheng et al., 2012) (Abbasi et al., 2012) (Pant & Srinivasan, 2013) (Jabr et al., 2013) (Bardhan et al., 2015) (Pant & Sheng, 2015)	Web, Predictive Analytics.
Cluster 2:	(Nelson et al., 2005) (Greenwald et al., 2010) (Susarla et al., 2010) (Koh et al., 2011) (Chau & Xu, 2012) (Park et al., 2012) (Wakefield, 2013) (Xixi Li et al., 2013) (Habjan et al., 2014) (Dong et al., 2014) (Popovič et al., 2014) (Raghunathan & Sarkar, 2016)	Procurement, SaaS, Data warehousing, ERP, Importance of Information in an organization.
Cluster 3:	(Arnott & Pervan, 2005) (T. D. Clark et al., 2007) (Kayande et al., 2009) (Arnott & Pervan, 2012) (Arnott & Pervan, 2014)	Decision support systems.
Cluster 4:	(Rouibah & Ould-Ali, 2002) (Petrini & Pozzebon, 2009) (H. Chen et al., 2012) (Lau et al., 2012) (Audzeyeva & Hudson, 2016) (Shollo & Galliers, 2016) (Gholami et al., 2016)	Business intelligence.
Cluster 5:	(Mookerjee & Mannino, 1997) (Baker et al., 2009) (Deng & Chi, 2012) (Sundararajan et al., 2013) (Meyer et al., 2014) (Agarwal & Dhar, 2014) (X.-B. Li & Sarkar, 2014) (D. Q. Chen et al., 2015) (Constantiou & Kallinikos, 2015) (Newell & Marabelli, 2015) (Ghoshal et al., 2015) (Loebbecke & Picot, 2015) (R. Clarke, 2016) (Abbasi et al., 2016) (Fayard et al., 2016) (Müller et al., 2016)	Big data, Big data analytics, Decision making.
Cluster 6:	(H. M. Kim et al., 2007) (Xin Li et al., 2009) (Luftman & Zadeh, 2011) (Luftman et al., 2012) (D. Hu et al., 2012) (Choudhary & Vithayathil, 2013) (Luftman et al., 2013) (Provost et al., 2015) (Luftman et al., 2015)	IT management, Data sharing, Risks, Privacy.
Cluster 7:	(Zhu et al., 1997) (Fan et al., 2005) (Roussinov & Chau, 2008) (G. Kim et al., 2011) (Coltman et al., 2011) (Lacity et al., 2011) (Abbasi et al., 2015)	Firm performance, Business Processes.

Table 2.5. Clustering of documents with K-means clustering and underlying themes

As is evident in Table 2.5, K-means clustering groups themes in the data set with greater definition than the bibliographic coupling technique.

2.6.5 Basic text mining of article abstracts

Due to the absence of abstracts in two articles, I only use 66 article abstracts as inputs to various text analyses techniques.

Word frequency and word clouds for the most frequently occurring words in my data set—with a minimum frequency of 18—are shown in Figure 2.9. The physical sizes of the terms in the word cloud are proportionate to their respective frequencies.

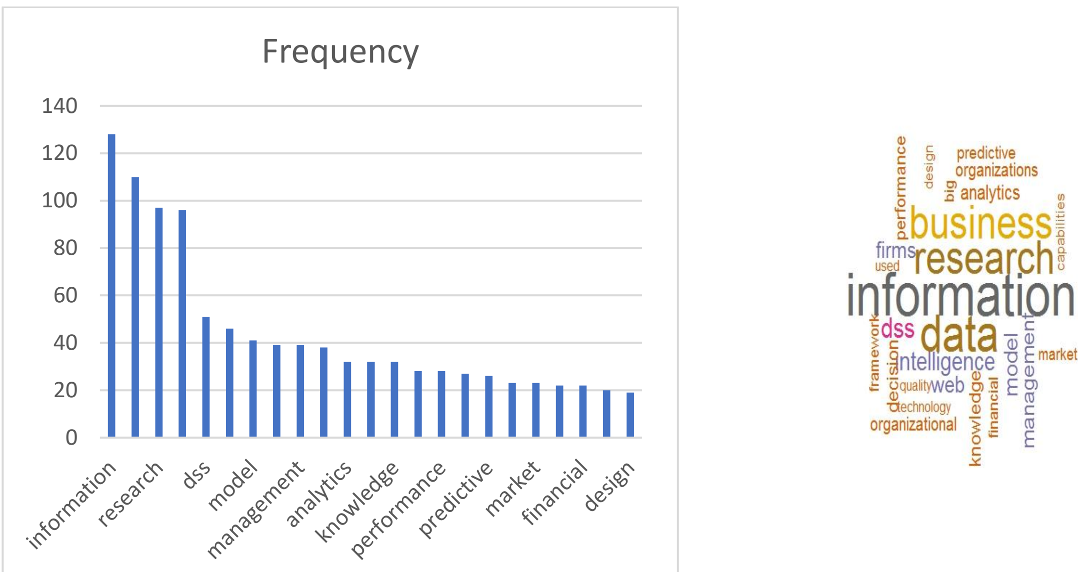


Figure 2.9. Frequency distributions and word cloud for frequently occurring terms (minimum frequency=18)

The dominance of keywords such as “information,” “data,” “research,” “business.” and “decision support system” in my data set highlights the importance of information for decision-making in BA. “Intelligence,” “model,” “analytics,” “decision,” and “performance” keywords suggest the importance of those concepts in achieving desired firm performance. Additionally, “web,” “market,” “financial,” “organizational,” “capabilities,” and “design” are important keywords from the BA domain.

My data set is temporally imbalanced in that it contains more articles published after 2010 than in the 13 previous years. Therefore, I cannot divide this corpus into equal time periods with an equal number of articles. Thus, I divided my corpus into three time periods by number of articles: 18 from 1997 to 2010; 22 from 2011 and 2013; and 26 from 2014 and 2016. Figure 2.10 shows word clouds with frequent keywords for these three time periods.

1997 to 2010 and 18 articles

2011 to 2013 and 22 articles

2014 to 2016 and 26 articles

Figure 2.10. Word clouds over three time periods (1997–2010; 2011–2013; and 2014–2016)

“Information,” “data,” and “business” are dominant words in all three time periods. Those frequencies suggest the importance of those topics in BA IS research both historically and contemporaneously. “Data” and “business” became more frequent keywords after 2010. “Web,” “markets,” “decision support,” “Decision Support System (DSS)”, “knowledge,” “intelligence,” and “process” are important words before 2010. “Web,” “predictive,” “intelligence,” “financial,” “capabilities,” “management,” “model,” Enterprise

Resource Planning (ERP)", "Business Process Outsourcing (BPO)," and "analytics" are words frequently mentioned in article abstracts between 2011 and 2013. Finally, "business data analytics (BDA)," "analytics," "organizational," "decision," "value," "big," "support," "decision," "analysis," "firms," "value," and "systems" are frequent words used in publications within the last three years. DSS and "decision support" were more popular keywords before 2010 became less popular during the next time period (2011–2013) but again became popular between 2014 and 2016. Similarly, "analytics" as a keyword was less popular before 2014 but subsequently gained popularity. "Financial," "capabilities," "ERP," "predictive," and "BPO" are used more frequently between 2011 and 2013 but did not occur after 2013. "Intelligence" or "business intelligence," popular buzzwords between 2011 and 2013, also lost popularity to "BDA" in later time periods.

Although the number of articles published in each of the journals in my data set differs per journal, I compare journals based on frequently occurring keywords and word clouds. Again, "information" and "data" are the most frequently occurring words across journals. "DSS" and "decision" are frequently used words in *EJIS*, *ISR*, *JIT*, and *MISQ*; "BDA" is prevalent in *EJIS* and *JMIS*; "intelligence" in *MISQ* and *JSIS*; "big data" in *ISJ* and *JAIS*; "performance" in *JAIS* and *EJIS*; and "analytics" is most frequently used only in *MISQ* articles.

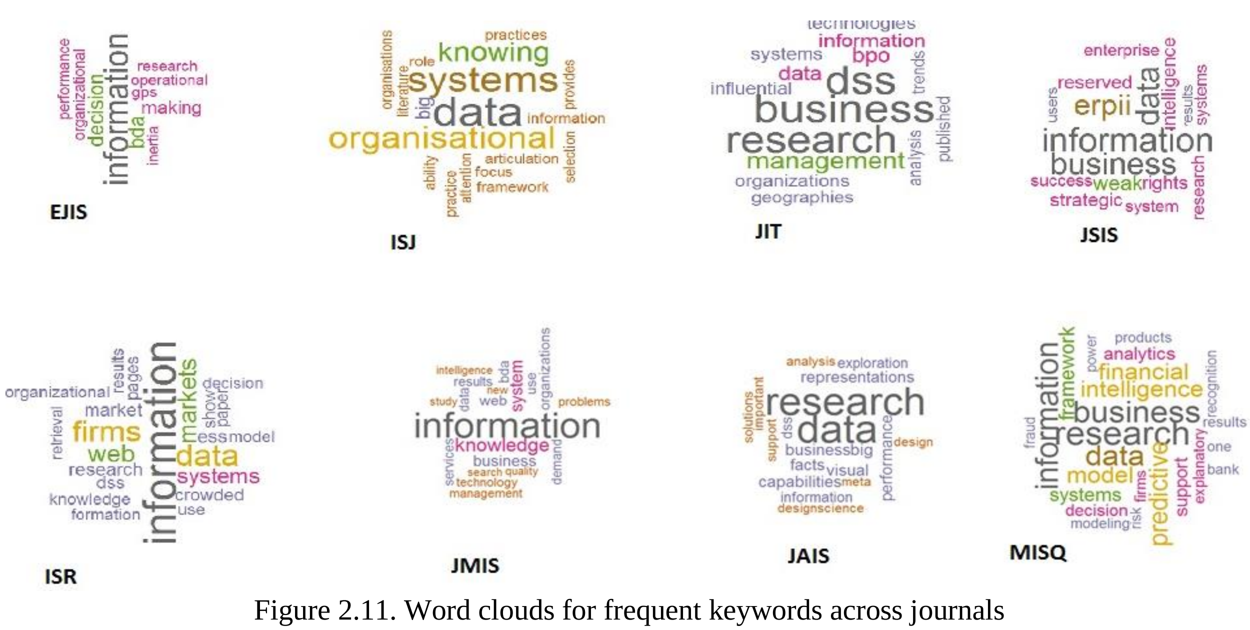


Figure 2.11. Word clouds for frequent keywords across journals

Word clouds for each journal prove that certain keywords are unique to specific journals. For example, Web," "Expert Support Systems (ESS)," "crowded," and "markets" are terms unique to *ISR* articles; "ERP" is unique to *JSIS*; "GPS" to *EJIS*; "predictive analytics" to *MISQ*; "BPO" to "JIT"; and "strategic" to *JSIS*. These keywords also suggest the topics in which these journals have been historically invested and might provide prospective BA IS researchers guidelines on potential targets for publication.

Finally, I compared journals for frequently occurring keywords by combining them into two groups: North American (*MISQ*, *ISR*, *JAIS*, and *JMIS*) and European (*EJIS*, *JSIS*, *JIT*, and *ISJ*). Word clouds for the two groups are shown in Figure 2.12.

"Information," "business," "research," "systems," "decision," and "data" are the most frequently occurring words in both groups, so they emphasize the topic areas across the BA IS research domain. "Web," "markets," "analytics," "model," and "predictive" are keywords prevalent in American journals, whereas "DSS," "BPO," "ERP," "analysis," and "process" are prevalent European journals.

Figure 2.12. Word cloud for North American (left) versus European Journals (right)

2.6.6 Determining topics with topic modeling

Finally, I used Latent Dirichlet Allocation (LDA), as implemented in MALLET (Machine Learning for Language Toolkit),[15] to uncover seven topics. I limited the program to seven topics in order to be consistent with the cluster analysis noted above. Table 2.6 illustrates key topics, associated words, and their top five corresponding articles.

[15] McCallum, Andrew Kachites. "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu. 2002.

Topic No.	Key Phrases	Top Five Articles	Underlying Theme
Topic 1:	information, web, status, pages, market, demand, recognition, mechanism, page, behavior, policy, revelation, show, crawlers, topical, products, link, sellers, locality, price	(Pant & Srinivasan, 2013) (Greenwald et al., 2010) (Dong et al., 2014) (Pant & Srinivasan, 2010) (Raghunathan & Sarkar, 2016)	Web, Predictive Analytics.
Topic 2:	predictive, firms, markets, analytics, ess, model, online, crowded, models, competitive, firm, find, explanatory, competing, performance, data, market, power, components, digital	(Chellappa et al., 2010) (Bardhan et al., 2015) (Shmueli & Koppius, 2011) (Pant & Sheng, 2015) (Zheng et al., 2012)	Predictive Analytics, Competitive Intelligence.
Topic 3:	Data, business, services, analytics, visual, rules, results, similar, privacy, quality, users, representations, facts, tasks, variety, made, context, theory, exploration, paper	(Baker et al., 2009) (Provost et al., 2015) (Newell & Marabelli, 2015) (Susarla et al., 2010), (H. M. Kim et al., 2007)	Privacy, Decision making.
Topic 4:	information, data, system, business, big, intelligence, research, organizational, study, organizations, bda, usage, based, systems, social, support, quality, strategic, framework, problems	(Xixi Li et al., 2013) (Fayard et al., 2016) (Popovič et al., 2014), (Deng & Chi, 2012) (D. Q. Chen et al., 2015)	IS Usage, Value creation.
Topic 5:	research, dss, systems, decision, model, support, design, analysis, bpo, empirical, business, science, major, making, management, information, published, decisions, journals, technology	(Arnott & Pervan, 2005) (T. D. Clark et al., 2007) (Kayande et al., 2009) (Arnott & Pervan, 2014), (Arnott & Pervan, 2012)	Decision support systems.
Topic 6:	business, management, organizational, based, research, paper, trends, firm, enterprise, capabilities, process, erpii, cloud, technologies, influential, geographies, survey, concerns, intelligence, factors	(Luftman et al., 2013) (Luftman & Zadeh, 2011) (Luftman et al., 2012) (Choudhary & Vithayathil, 2013) (Coltman et al., 2011)	IT management, firm performance.
Topic 7:	approach, web, information, based, retrieval, knowledge, financial, results, method, methods, proposed, framework, existing, phishing, study, performance, user, research, formation, bank	(Abbasi et al., 2015) (Mookerjee & Mannino, 1997), (Chung et al., 2005), (Abbasi et al., 2012), (Fan et al., 2005)	Web, Predictive Analytics.

Table 2.6 Topic modeling: Keywords, top articles, and underlying themes

As with common research themes obtained from clustering articles based on bibliographical coupling and K-means clustering, topic modeling suggests the following common themes in BA IS research: web, predictive analytics, IT management, business intelligence, IS usage, firm performance, and decision support.

2.6.6 Termite Plot

Termite plots are effective visualization techniques for assessing textual topic models with the intention of assessing the quality of individual topics (Chuang, Manning, & Heer, 2012). They can be viewed as term-topic matrices with topics represented in columns and frequently occurring terms in rows. Circular areas at the intersections of terms and topics show term probabilities on respective topics. The termite plot with seven topics and 25 frequently occurring terms from our corpus is shown in Figure 2.13.

Based on the probabilities (represented as circular areas) for the terms, few terms have very high probability loadings on certain topics. For example, Topic 1 concerns the web and markets; Topic 3 concerns predictive analysis and its use; Topic 4 concerns decision and decision support systems; Topic 5 concerns data and information use in the organization; and Topic 6 concerns BI's use in the organization. Topic 0 and Topic 7 share keywords and thus have equal probability loadings. Both topics are very general in that they concern data, information, analytics, and their use in firms or organizations.

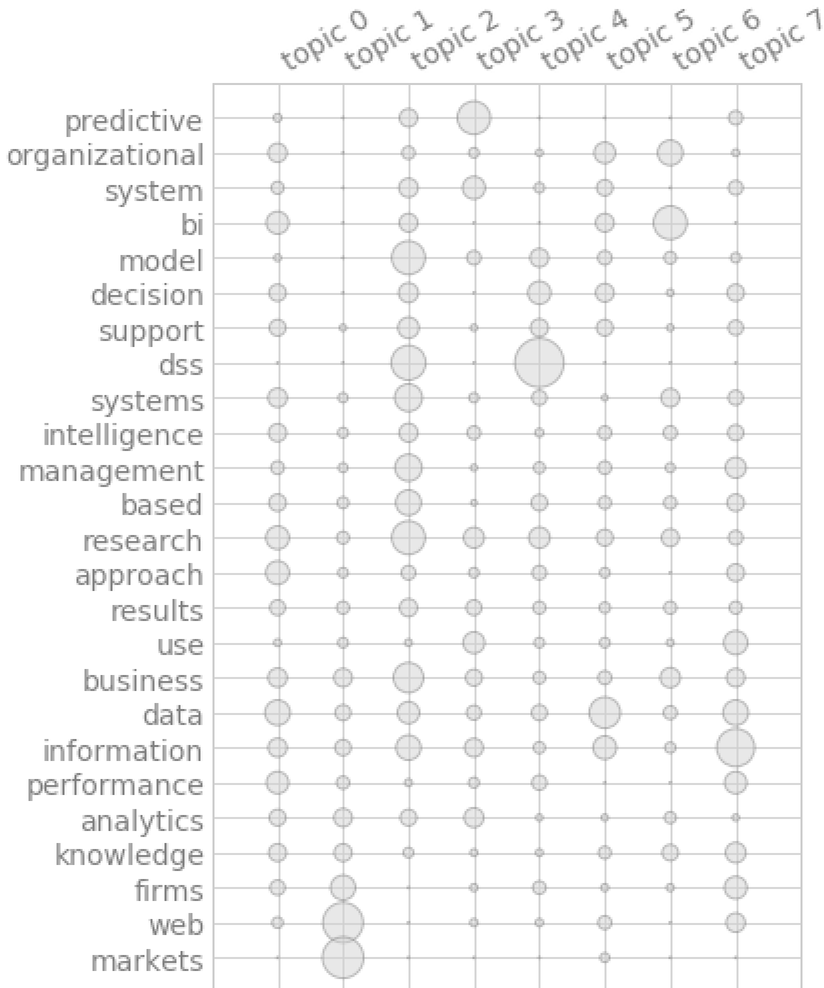


Figure 2.13. Termite Plot

2.6.7 Social Network Analysis for the co-authorship network

Co-authorship networks are a class of bibliometric networks that help study and analyze the social structure of a research community. Using ISI's Web of Science citation database, I performed social network analysis of the co-authorship network of 72 articles published 1997–2016 in leading IS journals with the same keywords mentioned in the data collection section. The data were collected in January 2017. Figure 2.14 shows the distribution of number of articles against number of authors across the journals.

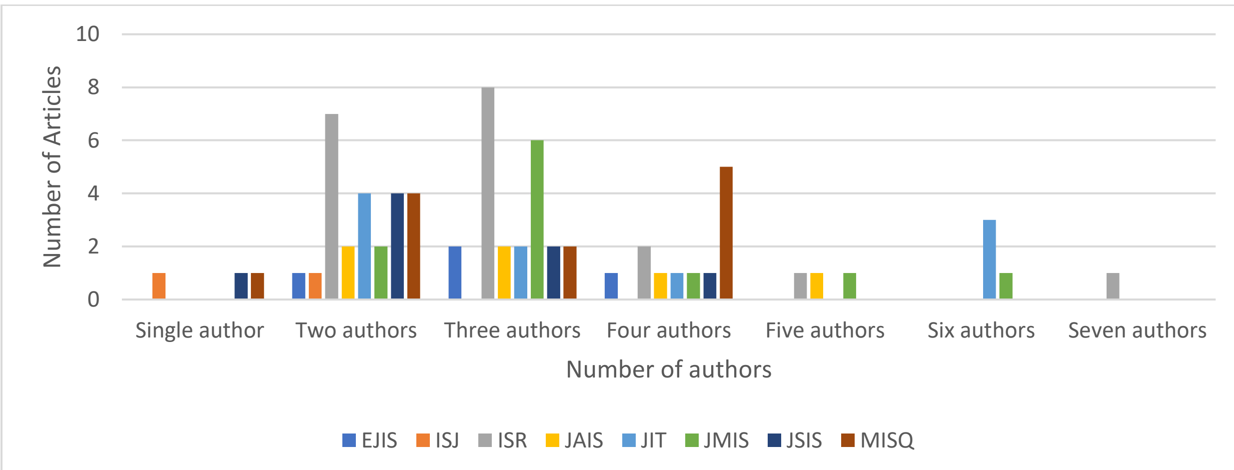


Figure 2.14. Distribution of authors across journals

The 72 publications are authored by 221 IS researchers, out of which 187 are unique. The data set contains three single-authored articles (1 each in *JSIS* and *ISJ* and 1 *MISQ* editorial), which I remove from the data set for this section. The resulting network of 184 unique authors and 148 edges is created using the statnet package in R statistical programming language and is shown in Figure 15. The network density is 0.008791, the mean degree is 1.6087, and standard deviation for degree distribution is 1.5218.

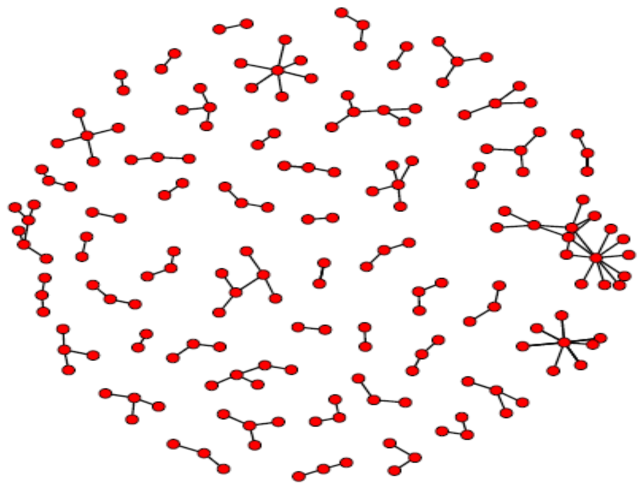


Figure 2.15. Co-authorship network with 184 unique authors

There are 51 components in our network with the largest component size 16 and the smallest size 2. Figure 16 shows the largest component, which has 16 vertices, 18 edges, and a network density of 0.15.

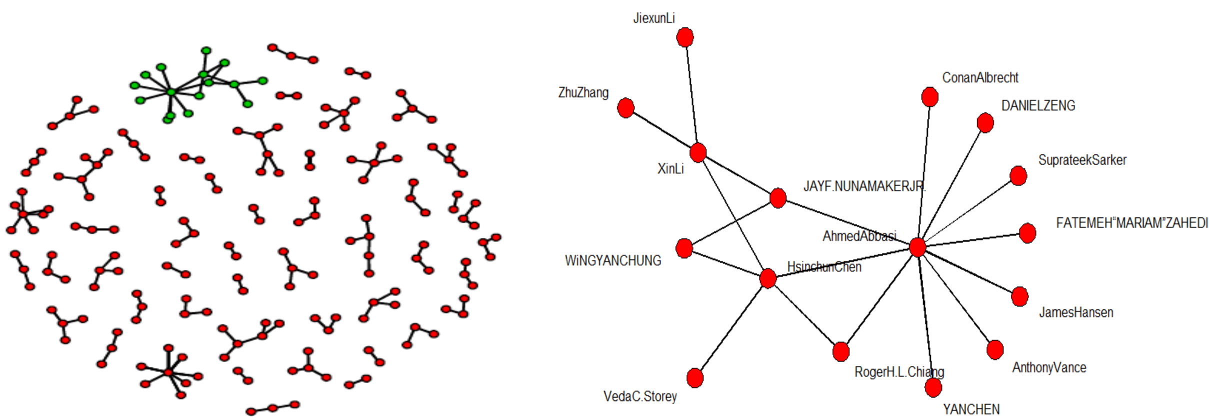


Figure 2.16. Largest component in co-authorship network (n=16, e=18, d=0.15)

Since my network is limited, various centrality measures like degree, betweenness, or eigenvalue centrality provide few insights and hence are not carried out in detail.

The degree distribution for the above co-authorship network is shown in Figure 2.17.

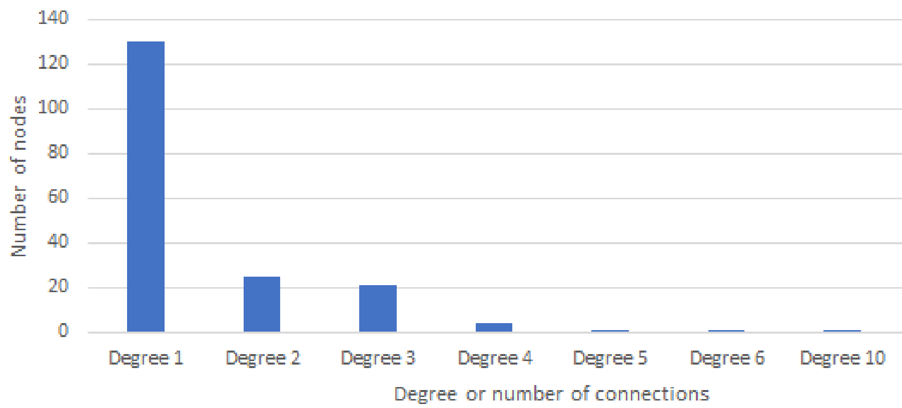


Figure 2.17. Degree distribution for co-authorship network

Since the distribution of relationships and the degree distribution for nodes (authors) in my network are not uniform, these relationships are not random. Certain authors collaborate more with one another than with others, and these propensities suggest that selection forces and/or author attributes shape the global structure of this network (Hunter et al., 2008a). These inclinations are evident where the ERGM social network analysis technique is applied to uses attributes of network members (gender, affiliations, etc.) to predict network structures in binary networks (Harris, 2013a).

Based on (Harris, 2013a), ERGM application for network statistical modeling begins with the intention of proving that the observed network is different from a random network with similar network characteristics (number of nodes, density, and edges). Therefore, I created a random network with a density of 0.00897, which resulted in a network with 184 vertices and 151 edges. The generated random network is shown in Figure 2.18.

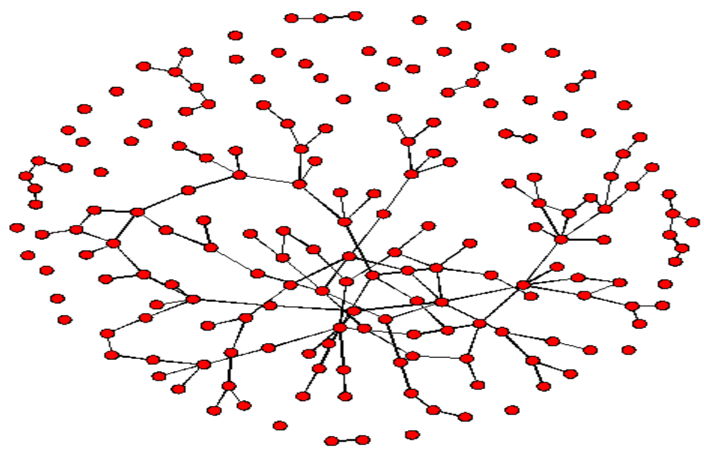


Figure 2.18. Random network with the same network characteristics as observed network

The mean degree for this random network is 3.2826 and standard deviation for degree distributions is 2.8219, both of which figures provide initial evidence of its distinction from the observed co-authorship network. The same difference is also evident when I compare the two networks for degree distributions.

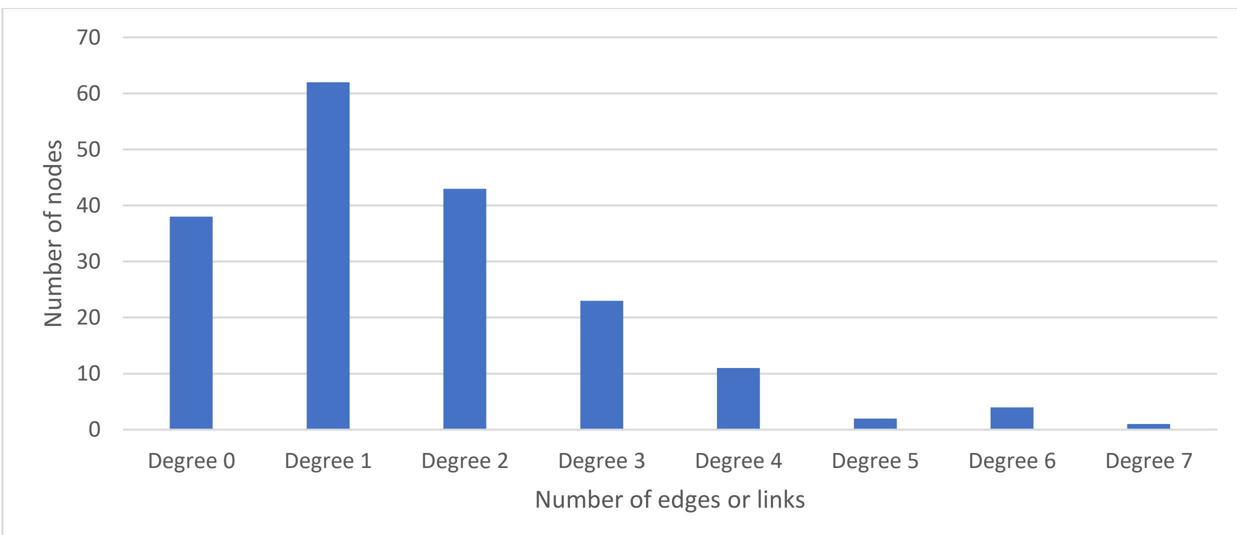


Figure 2.19. Degree distribution for a random network with similar network characteristics as observed co-authorship network

In addition to the degree distribution and visualizations, I compared the distribution of triads or subgraph counts with 16 different categories of possible triads for the observed network and the random network to pinpoint any transitivity biases in the observed network (Morris, Handcock, & Hunter, 2008). The triad distribution for the two networks is shown in Table 2.7 below.

	Observed Network	Random Network
No edges	996811	994179
One edge	24395	26930
Two edges	177	273
Three edges	1	2

Table 2.7 Comparing observed network with random network based on triad census

The triad census table shows that there are 996811 triads in the observed network with no edges, 24395 with one edge, 177 with two edges and only 1 complete triangle. Similarly, in the random network, there are 994179 triads with no edges, 26930 with one edge, 273 with two edges and 2 complete triangles. Thus, only slight difference exists between the two networks with respect to their network structures.

Finally, we plot distributions for shared partnerships in the form of dyadwise shared partners and edgewise shared partners as suggested in (Harris, 2013a; Morris et al., 2008) to further observe any differences in the two networks. Figure 2.20 shows resulting plots comparing observed with the random network.

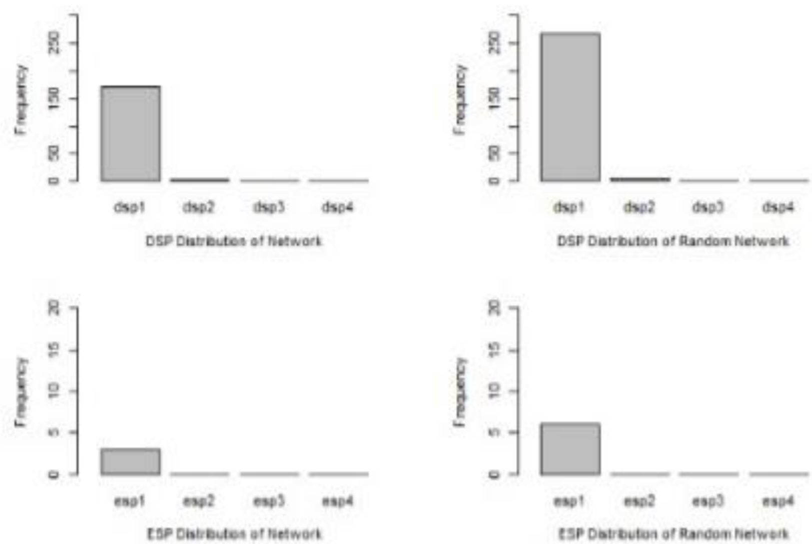


Figure 2.20. Plots comparing shared partnerships (DSP and ESP) for observed and random network

The above network visualizations exhibit the presence of potential clustering for the observed network based on underlying node attributes. Next, we check for the presence of any node clustering based on author attributes.

I collected the following author attributes to test for homophily: gender, continental affiliation (North American versus non-North American), university and departmental affiliation at publication time, current Carnegie classification for the affiliated university, Ph.D. completion year (as a proxy for academic experience), and presence or absence at a Ph.D.-granting university/institution. I analyzed the co-authorship network using the ERGM social network analysis technique to determine whether like author attributes are relevant antecedents for the formation of co-authorship network. Note that the present Carnegie classification of authors' affiliated universities is one of the independent variables to consider regarding whether the research reputation of the university is a factor in scholarly collaborations. Due to that variable, I acknowledge two complications in my study: 1) the Carnegie classification does not apply to European universities; and 2) the current Carnegie classification of universities may not reflect the research reputation of the institute when the article was published. Therefore, I tested the homophily model both with and without the Carnegie classification variable. Also, all the independent variables are factorial except for the category measuring an author's experience level, which is continuous. Table 2.8 shows summary distributions for author attributes for independent variables.

Author characteristic name	Common values
Gender	Female (n=43), Male (n=141)
Continental affiliation	North American (n=111), Others (n=73)
University affiliation	City University of Hong Kong (n=7), New York University (n=7), the University of Texas at Dallas (n=7), University of Iowa (n=5), University of Minnesota (n=5), University of Virginia (n=5), Others (n=148)
Departmental affiliation	IS (n=119), CSE (n=8), Economics (n=8), IT (n=8), Marketing (n=8), Operations Management (n=8, Others (n=25)
Experience	Minimum = 0.00, Median = 17, Maximum = 54
Ph.D. granting institution	University of Arizona (n=9), the University of Texas at Austin (n=7), Carnegie Mellon University (n=6), New York University (n=5), University of Georgia (n=5), Others (n=152)
Current Carnegie classification of author's university affiliation	Doctoral Universities: Highest Research Activity (n=76) Doctoral Universities: Higher Research Activity (n=20) Doctoral Universities: Limited Research Activity (n=1) Master's Colleges & Universities: Larger Programs(n=5) Baccalaureate Colleges: Arts & Sciences Focus(n=1) Special Focus Four-Year: Medical Schools & Centers(n=1) NON-USA(n=80)

Table 2.8 Summary distribution for important author attributes

2.6.7.1 Statistical network modeling results

As with traditional forms of model building in empirical research, network modeling for the statistical network also begins with a null model consisting of a single term representing the edges in the network—i.e., n~edges. The MLE coefficient for the null model is negative (-4.8105) indicating that the density of the network is below 50 % or less than 0.5 (Harris, 2013a). The p-value for this coefficient is also highly statistically significant with a value of less than 10^{-4} indicating that the presence of an edge is a significant indicator of network formation. The summary result of the null model is in Appendix A as Figure A.1.

The probability of link formation as per logistic regression model for above edge estimate of -4.8105 is calculated using formula 1 / (1+e$^{-(-4.8105)}$) which is (e$^{-(-4.8105)}$ is 122.793) 0.008078 or 0.8078 % which is roughly same as the density of our network 0.008790687. The AIC and BIC values for the null model are 1584 and 1591 respectively.

Although the null model is a good representation of the density of our observed network, it is unlikely that it represents other network characteristics well. As suggested by (Goodreau, Handcock, Hunter, Butts, & Morris, 2008) when the plot of triangle distributions for 100 simulated networks based on the null model is compared with my observed network, I found major differences, as shown in Figure 2.21.

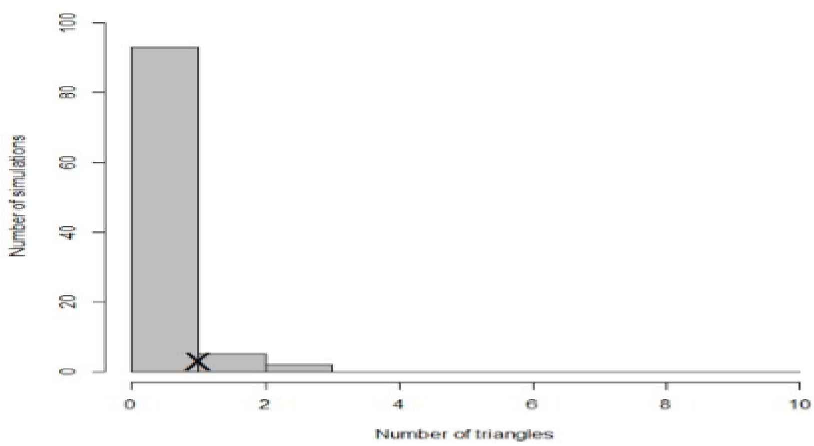


Figure 2.21. Number of triangles in 100 simulated networks based on the null model. X marks the number of triangles in the observed network

For the main effects model, I added all the independent variables with the node attributes of gender, continental affiliation, university and departmental affiliations, status at a Ph.D.-granting institution, experience (Ph.D. completion year), and current Carnegie classification. None of the independent variables is significant predictor of link formation for this study's observed co-authorship network for the main effects model. The AIC, BIC values for this model are 1957 and 3774. When I did not include the current Carnegie classifications in the main model, authors' affiliations with Ph.D.-granting institutions become significant (especially, University of Cincinnati and University of Georgia) as did university affiliations (especially, University of Hong Kong and University of Iowa). These implications, which suggest homophily based on these two variables, must be investigated further. The AIC and BIC values for this model are 1945 and 3716, respectively. From the AIC, BIC values, the main effects model without current Carnegie classification is more relevant than the model that includes this variable. The summary results of these two models are shown in Figure A.2 and A.3 of Appendix A.

The homophily model is specified with the “nodematch” attribute of the R statistical language’s “ergm” package (Hunter et al., 2008a; Morris et al., 2008). The homophily models are run both with and without the inclusion of the factorial variable representing current Carnegie classification. The results of the homophily model are shown in Figures A.4 (with current Carnegie classification) and A.5 (without current Carnegie classification) in Appendix A. Figure 2.22 shows the results of the null and two variants of homophily models.

ERGM results for Null model and Homophily models

	Dependent variable:		
	null model (1)	n homophily model with Carnegie (2)	homophily model without Carnegie (3)
edges	-4.811*** (0.086)	-5.792*** (0.227)	-5.762*** (0.228)
nodematch.Gender		0.158 (0.190)	0.165 (0.190)
nodematch.National_affiliation		0.163 (0.238)	0.388** (0.196)
nodematch.Dept_affiliation		0.724*** (0.183)	0.727*** (0.182)
nodematch.Experience		-0.188 (0.472)	-0.152 (0.472)
nodematch.PhD_granting_institute		1.910*** (0.291)	1.906*** (0.290)
nodematch.University_affiliation		2.638*** (0.275)	2.745*** (0.270)
nodematch.Carnegie_classification		0.393* (0.225)	
Akaike Inf. Crit.	1,583.563	1,446.642	1,447.754
Bayesian Inf. Crit.	1,591.294	1,508.492	1,501.873
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 2.22. ERGM results for null and homophily models for co-authorship network

The results are similar for the other independent variables in most cases. However, in the homophily model with Carnegie classification, continental affiliation is insignificant (p=0.4918) and Carnegie classification is significant (p=0.0808). For the homophily based on departmental affiliation, Ph.D.-granting institution and university affiliation are significant for both homophily models. The coefficients for gender and experience are statistically insignificant. Thus, three forms of affiliation (continental, university, and departmental) along with homophily based on Ph.D.-granting institution are significant predictors of network formation. The AIC, BIC values for the homophily model which includes Carnegie classification are 1447 and 1508, respectively. The AIC and BIC values for the homophily model that does not include the Carnegie classification as an independent variable are 1448 and 1502, respectively. Comparing these model fit values with those of the null model whose AIC and BIC values are 1584 and 1591, respectively (with only edges as significant predictors of network formation), shows that both observed homophily models are more relevant than the null model.

The results from the homophily model without the Carnegie classification variable provide interesting insights. Lack of homophily based on gender shows an increasing number of female researchers in BA IS (n=43). For homophily based on continental affiliation and department affiliation, results are as hypothesized. That is, most authors affiliated with the US universities tend to work with other US authors—and vice versa. Results show homophily based on departmental affiliation perhaps due to most authors having been affiliated with the information systems department (n=119). The study is also conducted on leading IS journals which may not appeal to non-IS researchers. The insignificance of homophily based on experience may exist due to the prevalence of research partnerships wherein senior scholars (mostly professors) work with and guide junior faculty and Ph.D. students towards co-authored publications in the renowned information systems journals. The most interesting insight is the significance of homophily based on Ph.D.-granting institution, which suggests that even after completion of the doctoral degree, researchers continue collaborating for scholarly publications. One possible reason for this result apart from graduating from the same program might be the brokerage role played by Ph.D. supervisors or supervising committee members, bringing together scholar alumni for research collaborations. Unsurprisingly, homophily based on university affiliation is also significant due to the importance of working in the geographical vicinity on university-wide, interdisciplinary projects.

Thus, the homophily model can successfully estimate the effect of certain attributes of nodes on link formation. In addition, homophily is measurable for each level of an individual categorical variable with differential homophily models (Harris, 2013a). In differential homophily, homophily terms are included separately for each node factor of each categorical variable in the model with “diff=true” attribute. The output from the differential homophily models (both with and without the variable representing Carnegie classification) is available in Appendix A, Figures A.6 and A.7, respectively. Note that most of the coefficients for an independent variable are estimated as negative infinity. These results occurred because my data set contains few authors with the desired values for these variables, and hence they cannot form network ties. Based on a comparison of the AIC and BIC values for the differential homophily models and homophily models, one can argue that the homophily models are slightly more relevant than their differential homophily variants.

2.6.7.2 Goodness of fit of homophily model

Based on (Harris, 2013a), one can also use simulation to assess model fit by simulating a single network based on a model and comparing the network characteristics of the simulated network with the network characteristics of the observed network. The simulation results for our most efficient models (null and homophily) are shown in Table 2.9.

	edges	degree0	degree1	degree2	degree3	degree4	degree5	degree6	degree7	degree8	degree9	degree10	triangle
Observed network	136	0	138	23	16	3	1	1	1	0	0	1	1
Null model	140	43	53	53	24	6	5	0	0	0	0	0	1
Homophily model without Carnegie	149	40	59	41	24	16	3	1	0	0	0	0	4
Homophily model with Carnegie	128	50	59	43	21	7	4	0	0	0	0	0	2

Table 2.9. Simulation networks for null and homophily models to test for goodness of fit over degree distributions and triangles

Although the null model appears to accord more closely with results charted in the simulation network, researchers need to undertake further tests of goodness of fit to determine which models are the most relevant. Currently. all of the aforementioned models fit too closely to the observed network.

Goodness-of-fit procedures can compare certain network characteristics such as degree, edge-wise shared partners (ESP), and dyad-wise shared partners (DSP) for simulated network models and the observed network (Harris, 2013a). These comparisons are built into a goodness of fit procedure within the “ergm” package in R language. Hence, observed frequencies for these network characteristics can be compared with frequencies in the simulated models. These comparisons include observed, minimum value, mean value, maximum value, and MC p-value for 100 simulations of null. One of the homophily models is shown in Figure 2.23.

```
Goodness-of-fit for degree

   obs min  mean max MC p-value
0    0  24 42.05  60       0.00
1  138  44 62.02  74       0.00
2   23  32 45.97  60       0.00
3   16  10 22.18  37       0.34
4    3   2  8.65  17       0.14
5    1   0  2.25   7       0.74
6    1   0  0.65   4       0.96
7    1   0  0.20   2       0.38
8    0   0  0.02   1       1.00
9    0   0  0.01   1       1.00
10   1   0  0.00   0       0.00

Goodness-of-fit for edgewise shared partner

     obs min   mean max MC p-value
esp0 133 111 134.20 165       0.98
esp1   3   0   1.73  14       0.76
esp2   0   0   0.02   2       1.00

Goodness-of-fit for dyadwise shared partner

       obs   min     mean   max MC p-value
dsp0 16661 16525 16635.31 16710       0.60
dsp1   171   126   199.61   305       0.56
dsp2     3     0     1.07     6       0.30
dsp3     1     0     0.01     1       0.02
```

```
Goodness-of-fit for degree

   obs min  mean max MC p-value
0    0  25 41.32  55       0.00
1  138  46 61.73  79       0.00
2   23  32 46.55  64       0.00
3   16  13 22.71  34       0.16
4    3   3  8.32  17       0.08
5    1   0  2.58   8       0.60
6    1   0  0.62   4       0.90
7    1   0  0.12   1       0.24
8    0   0  0.03   1       1.00
9    0   0  0.02   1       1.00
10   1   0  0.00   0       0.00

Goodness-of-fit for edgewise shared partner

     obs min   mean max MC p-value
esp0 133 109 135.69 168       0.84
esp1   3   0   1.36   9       0.68
esp2   0   0   0.01   1       1.00

Goodness-of-fit for dyadwise shared partner

       obs   min     mean   max MC p-value
dsp0 16661 16530 16633.37 16719       0.44
dsp1   171   117   201.51   306       0.42
dsp2     3     0     1.09     5       0.24
dsp3     1     0     0.03     1       0.06
```

(a). The goodness of fit for null model (b). The goodness of fit for homophily model

Figure 2.23. The goodness of fit for null versus homophily models for degree distribution, ESP (edgewise shared partners) and DSP (dyad-wise shared partners)

Large MC p-values are indicators of a simulated network that is similar (i.e., not significantly different) from the observed network for a network characteristic. In contrast, small MC p-values are indicators of differences between observed and simulated frequencies for specific network characteristics. P-values less than 0.05 indicate significant differences between simulated model networks and observed networks. In such cases, researchers can conclude that that the model is not relevant to the data (Harris, 2013a).

Based on observations for p-values above, the homophily model outperforms the null model for edgewise shared partnerships, but it underperforms for the goodness of fit for degree distributions and for all values for dyad-wise shared partnerships except for one (DSP3).

To further investigate the results from the two variants of the homophily models, I carried out Poisson regression with counts of degree as the dependent variable and the author attributes as independent variables. As before, the Poisson regression is performed both with and without the present Carnegie classification. The results from Poisson regression are similar to those obtained from the homophily models with ERGM, except continental affiliation is insignificant in both models. From Poisson regression, I also learned the specific individual attribute values that are statistically significant. For example, the university affiliations attribute is statistically significant for researchers from Loisisinia State University, University of Virginia, University of Iowa, and a few others. Similarly, Economics, IS, and IT field affiliations have a statistically significant relationship with a number of degrees or author collaborations. Finally, Ph.D.. graduates from University of Arizona, University of Minnesota, and University of Pittsburg tend to publish together in leading information systems journals even after graduate students complete their degrees. . These results are similar to those from the differential homophily model, which also allows checking for the statistical significance of each attribute value of the individual categorical variables. The results of the Poisson regression are shown in Appendix A, Figure A.8 (with Carnegie classification) and Figure A.9 (without Carnegie Classification).

2.7 Discussion

This study expounds upon and complicates business analytics research in the information systems disciplines by examining specific and systemic relationships between ideas, authors, and the public and private institutions in which BA IS research takes place. First, this chapter identifies common reference disciplines in BA IS, which knowledge also contributes to scholars' understandings of disciplinary boundaries for BA IS research. These results provide guidelines for information systems researchers working in BA regarding target journals for certain topics of publication. The second theoretical contribution of this study is the detection of author and article clusters that each focus on specific field subtopics. Thus, my work contributes to the existing work of community detection in scholarly networks and in the detection of underlying topics from such networks (Ding, 2011; Yan, Ding, Milojević, & Sugimoto, 2012). I was also able to employ text analysis of topic modeling to detect other underlying topics. The third contribution of this study is the comparison of journals and journal groups, which are based on two factors: reference disciplines and keywords. These factors allowed me to discern topic inclinations and social network affiliations for certain journals and journal groups.

Based on the social network analysis of co-authorship network, the main theoretical contribution of this chapter is the apparent effectiveness of implementing the underutilized technique of Exponential Random Graph Modeling (ERGM) in IS research. I used ERGM to test for the homophily principle for the co-authorship network of authors for leading BA IS journals. Although there were few extant IS studies using ERGM (Faraj & Johnson, 2011; Shi, Lee, & Whinston, 2015), researchers can employ this technique for a variety of research scenarios based on social networks in the IS domain, such as healthcare networks, online communities, and organizational networks.

2.8 Conclusion

Defining the intellectual structure of a discipline helps scholars and practitioners design research projects by understanding the historical development of the field and by demarcating its boundaries with other related disciplines. This study helps distinguish the domain of business analytics research as carried out by IS researchers when publishing in elite IS journals. Identification of the references disciplines are indicators of the influence of other disciplines on BA IS research, and topics extracted from the articles' abstracts are indicators of important research areas. Inter-citation patterns exhibit nuanced differences between journals and journal groups. This chapter identifies 11 reference disciplines, seven topics, and several author attributes as significant predictors of the co-authorship network. Conclusions regarding target journals, keywords, and BA homophily models will help future scholars define research projects, target journals, and perhaps even choose university and/or continental affiliations.

Appendix A

```
==========================
Summary of model fit
==========================

Formula:   n ~ edges

Iterations:  7 out of 20

Monte Carlo MLE Results:
      Estimate Std. Error MCMC % p-value
edges  -4.8105     0.0861      0  <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 23340  on 16836  degrees of freedom
 Residual Deviance:  1582  on 16835  degrees of freedom

AIC: 1584    BIC: 1591    (Smaller is better.)
```

Figure A.1. Summary of null model

```
==========================
Summary of model fit
==========================

Formula:   n ~ edges + nodefactor("Gender") + nodefactor("National_affiliation") +
    nodefactor("Dept_affiliation") + nodecov("Experience") +
    nodefactor("Carnegie_classification") + nodefactor("PhD_granting_institute") +
    nodefactor("University_affiliation")

Iterations:  8 out of 20

Monte Carlo MLE Results:
                                                                  Estimate Std. Error MCMC % p-value
edges                                                            -4.775615   4.499290      0  0.2885
nodefactor.Gender.Male                                            0.526548   0.641226      0  0.4116
nodefactor.National_affiliation.Non_american                     -1.295452   3.831427      0  0.7353
nodefactor.Dept_affiliation.Decision Sciences                    -1.849206   5.045070      0  0.7140
nodefactor.Dept_affiliation.Economics                             2.232847   2.035362      0  0.2726
nodefactor.Dept_affiliation.Engineering Management                0.048697   1.666540      0  0.9767
nodefactor.Dept_affiliation.Finance                               0.048697   1.666540      0  0.9767
nodefactor.Dept_affiliation.IE                                    1.759680   3.392553      0  0.6040
nodefactor.Dept_affiliation.Industry                              1.357827   2.275848      0  0.5508
nodefactor.Dept_affiliation.Innovation and Entrepreneurship       0.071660   2.221895      0  0.9743
nodefactor.Dept_affiliation.International Business                0.565039   1.544871      0  0.7146
nodefactor.Dept_affiliation.IS                                    0.565039   0.601803      0  0.3478
nodefactor.Dept_affiliation.IT                                    0.448459   1.373345      0  0.7440
nodefactor.Dept_affiliation.IT Management                         0.544628   1.159252      0  0.6385
nodefactor.Dept_affiliation.Marketing                             0.588002   1.590372      0  0.7116
nodefactor.Dept_affiliation.OM                                    0.524783   1.095170      0  0.6318
nodefactor.Dept_affiliation.Psychology                           -0.673866   2.944870      0  0.8190
nodefactor.Dept_affiliation.Statistics                            0.919721   2.581734      0  0.7217
nodefactor.Dept_affiliation.Technology Management                -1.078801   3.149912      0  0.7320
nodecov.Experience                                               -0.002551   0.024839      0  0.9182
nodefactor.Carnegie_classification.16                             1.364549   3.325717      0  0.6816
nodefactor.Carnegie_classification.17                             1.662900   5.216572      0  0.7499
nodefactor.Carnegie_classification.18                             1.809310   2.303094      0  0.4321
nodefactor.Carnegie_classification.21                            -1.487730   2.644899      0  0.5738
```

```
nodefactor.Carnegie_classification.25                                       -2.198871   3.894885      0  0.5724
nodefactor.Carnegie_classification.99                                        0.347028   2.214398      0  0.8755
nodefactor.PhD_granting_institute.Australian Graduate School of Management   0.194180   1.386784      0  0.8886
nodefactor.PhD_granting_institute.Cambridge University                       0.076763   1.719282      0  0.9644
nodefactor.PhD_granting_institute.Cardiff University                         0.051028   1.507065      0  0.9730
nodefactor.PhD_granting_institute.Case Western Reserve University           -1.015682   1.701733      0  0.5506
nodefactor.PhD_granting_institute.Chinese University of HongKong            -0.352172   2.229883      0  0.8745
nodefactor.PhD_granting_institute.City University of Hong Kong              -0.868515   2.316371      0  0.7077
nodefactor.PhD_granting_institute.Clemson University                        -0.784861   2.545948      0  0.7579
nodefactor.PhD_granting_institute.CMU                                       -1.670160   3.191342      0  0.6007
nodefactor.PhD_granting_institute.Columbia University                       -1.421393   3.131590      0  0.6499
nodefactor.PhD_granting_institute.Copenhagen Business School                -0.243762   1.410502      0  0.8628
nodefactor.PhD_granting_institute.Cranfield University                      -0.399762   1.796806      0  0.8239
nodefactor.PhD_granting_institute.Curtin University                         -0.516343   1.556393      0  0.7401
nodefactor.PhD_granting_institute.Ecole des Hautes Etudes en Sciences Sociales  0.894320   3.196020      0  0.7796
nodefactor.PhD_granting_institute.Ecole Superiere des Affaires              -0.534202   1.567974      0  0.7333
nodefactor.PhD_granting_institute.Erasmus University                        -0.523997   1.558994      0  0.7368
nodefactor.PhD_granting_institute.Florida Inernational University           -1.452010   2.704729      0  0.5914
nodefactor.PhD_granting_institute.Florida International University          -0.508688   1.557357      0  0.7439
nodefactor.PhD_granting_institute.Florida State University                  -0.703216   2.743958      0  0.7977
nodefactor.PhD_granting_institute.Fundação Getúlio Vargas                   -0.007654   1.424785      0  0.9957
nodefactor.PhD_granting_institute.Georgia State University                  -0.531651   1.565143      0  0.7341
nodefactor.PhD_granting_institute.HKUST                                     -0.306027   2.380214      0  0.8977
nodefactor.PhD_granting_institute.HongKong Polytechnic University           -0.980248   2.305500      0  0.6707
nodefactor.PhD_granting_institute.Huazhong University of Science and Technology  -2.202010   2.650485      0  0.4061
nodefactor.PhD_granting_institute.IIT India                                 -0.415070   1.836428      0  0.8212
nodefactor.PhD_granting_institute.Indiana University                        -2.979144   3.112630      0  0.3385
nodefactor.PhD_granting_institute.Institute of France Telecom               -0.853206   2.329841      0  0.7142
nodefactor.PhD_granting_institute.Johns Hopkins University                  -0.447171   2.018686      0  0.8247
nodefactor.PhD_granting_institute.KAIST Business School                     -0.891477   1.824971      0  0.6252
nodefactor.PhD_granting_institute.Katholieke University Leuven                     NA   0.000000      0      NA
nodefactor.PhD_granting_institute.Kent State University                     -2.919418   2.802070      0  0.2975
nodefactor.PhD_granting_institute.Kuwait University                          0.010206   1.426300      0  0.9943
nodefactor.PhD_granting_institute.LSE                                       -3.613197   4.127835      0  0.3814
```

```
nodefactor.PhD_granting_institute.McGill University                                NA   0.000000      0      NA
nodefactor.PhD_granting_institute.Medical College of Wisconsin              -0.540162   2.101557      0  0.7972
nodefactor.PhD_granting_institute.MIT                                       -1.768939   1.804344      0  0.3269
nodefactor.PhD_granting_institute.Monash University                         -0.516343   1.556393      0  0.7401
nodefactor.PhD_granting_institute.NA                                         0.750779   1.601909      0  0.6393
nodefactor.PhD_granting_institute.Northwestern University                   -1.452010   3.056145      0  0.6347
nodefactor.PhD_granting_institute.Nova Information Management School        -0.871024   2.837016      0  0.7588
nodefactor.PhD_granting_institute.NUS                                       -1.342298   2.096081      0  0.5219
nodefactor.PhD_granting_institute.NYU                                       -0.531587   1.502243      0  0.7234
nodefactor.PhD_granting_institute.Pennsylvinia State University             -2.937848   3.161439      0  0.3528
nodefactor.PhD_granting_institute.Purdue University                         -0.667390   1.648011      0  0.6855
nodefactor.PhD_granting_institute.Queensland University of Technology        0.244460   2.163747      0  0.9100
nodefactor.PhD_granting_institute.RMIT                                      -0.506137   1.858339      0  0.7853
nodefactor.PhD_granting_institute.Stevens Institute of Technology           -0.767817   4.998087      0  0.8779
nodefactor.PhD_granting_institute.Syracuse University                       -1.777372   2.534482      0  0.4831
nodefactor.PhD_granting_institute.Technion Israel Institute of Technology          NA   0.000000      0      NA
nodefactor.PhD_granting_institute.Texas Tech                                 0.165895   1.949005      0  0.9322
nodefactor.PhD_granting_institute.UBC                                       -3.194698   2.293571      0  0.1637
nodefactor.PhD_granting_institute.UC Berkeley                               -0.850655   2.333005      0  0.7154
nodefactor.PhD_granting_institute.UC Irvine                                 -1.500487   2.634349      0  0.5690
nodefactor.PhD_granting_institute.UCLA                                      -0.498483   1.564173      0  0.7500
nodefactor.PhD_granting_institute.Universidade do Vale do Rio dos Sinos     -0.516343   1.384285      0  0.7092
nodefactor.PhD_granting_institute.Universita Cattolica del Sacro Cuore      -2.004072   3.017958      0  0.5067
nodefactor.PhD_granting_institute.University of Alberta                      0.886336   1.934726      0  0.6469
nodefactor.PhD_granting_institute.University of Arizona                     -1.464767   2.269146      0  0.5186
nodefactor.PhD_granting_institute.University of Bradford                    -0.539305   2.122369      0  0.7994
nodefactor.PhD_granting_institute.University of Cambridge                    0.578772   1.965764      0  0.7684
nodefactor.PhD_granting_institute.University of Cape Town                   -0.516343   1.556393      0  0.7401
nodefactor.PhD_granting_institute.University of Chicago                     -1.616779   1.901178      0  0.3951
nodefactor.PhD_granting_institute.University of Cincinnati                  -3.781235   2.576320      0  0.1422
nodefactor.PhD_granting_institute.University of Cologne                      1.664252   3.262245      0  0.6099
nodefactor.PhD_granting_institute.University of Georgia                     -3.050018   2.200888      0  0.1658
nodefactor.PhD_granting_institute.University of Houston                     -1.161727   4.669365      0  0.8035
nodefactor.PhD_granting_institute.University of Iowa                        -2.256684   2.971958      0  0.4477
nodefactor.PhD_granting_institute.University of Kentucky                    -2.604582   5.228782      0  0.6184
nodefactor.PhD_granting_institute.University of Leeds                              NA   0.000000      0      NA
nodefactor.PhD_granting_institute.University of Liechtenstein               -0.559717   1.617133      0  0.7293
nodefactor.PhD_granting_institute.University of Liverpool                   -0.541857   1.578784      0  0.7314
```

nodefactor.PhD_granting_institute.University of Ljublijana	-1.490343	2.492931	0	0.5500
nodefactor.PhD_granting_institute.University of Maryland	-2.781405	5.207145	0	0.5932
nodefactor.PhD_granting_institute.University of Michigan	-2.230714	1.942135	0	0.2507
nodefactor.PhD_granting_institute.University of Minnesota	0.977739	1.705987	0	0.5666
nodefactor.PhD_granting_institute.University of Missouri	-2.615920	4.767863	0	0.5832
nodefactor.PhD_granting_institute.University of Muenster	0.576221	1.361914	0	0.6722
nodefactor.PhD_granting_institute.University of Munich	-2.107608	2.641980	0	0.4250
nodefactor.PhD_granting_institute.University of Nottingham	-2.604582	5.228782	0	0.6184
nodefactor.PhD_granting_institute.University of Oklahama	-0.798675	2.422534	0	0.7416
nodefactor.PhD_granting_institute.University of Oulu	-1.778951	2.558830	0	0.4869
nodefactor.PhD_granting_institute.University of Pennsylvinia	-0.508907	1.705610	0	0.7654
nodefactor.PhD_granting_institute.University of Pittsburgh	-1.153221	1.802373	0	0.5223
nodefactor.PhD_granting_institute.University of Rochester	-0.688023	1.715365	0	0.6884
nodefactor.PhD_granting_institute.University of Sourthern California	-0.523997	1.558994	0	0.7368
nodefactor.PhD_granting_institute.University of South Carolina	-2.829316	4.825342	0	0.5577
nodefactor.PhD_granting_institute.University of Strathclyde	0.977696	3.056416	0	0.7491
nodefactor.PhD_granting_institute.University of Toronto	0.194180	1.386784	0	0.8886
nodefactor.PhD_granting_institute.University of Utah	-0.516343	1.556393	0	0.7401
nodefactor.PhD_granting_institute.University of Washington	-1.968839	2.156811	0	0.3613
nodefactor.PhD_granting_institute.University of Wisconsin Milwaukee	-2.302768	4.673565	0	0.6222
nodefactor.PhD_granting_institute.University of Wollongong	0.007654	1.424785	0	0.9957
nodefactor.PhD_granting_institute.Uppsala University	-0.480623	1.591029	0	0.7626
nodefactor.PhD_granting_institute.UT Arlington	NA	0.000000	0	NA
nodefactor.PhD_granting_institute.UT Austin	-0.488496	1.509675	0	0.7463
nodefactor.PhD_granting_institute.UT Dallas	-0.384254	2.112659	0	0.8557
nodefactor.PhD_granting_institute.Vrije University	-0.432930	1.896459	0	0.8194
nodefactor.PhD_granting_institute.Washington State University	NA	0.000000	0	NA
nodefactor.PhD_granting_institute.Yonsei University	0.606838	1.325048	0	0.6470
nodefactor.University_affiliation.AT&T Lab	-1.697946	4.213154	0	0.6869
nodefactor.University_affiliation.Auburn University	NA	0.000000	0	NA
nodefactor.University_affiliation.Australian National University	NA	0.000000	0	NA
nodefactor.University_affiliation.Bentley University	-1.298071	2.573933	0	0.6140
nodefactor.University_affiliation.Blue Slate Solutions	1.304027	2.639291	0	0.6213
nodefactor.University_affiliation.Brown University	0.863828	2.298192	0	0.7070
nodefactor.University_affiliation.Brunel University	NA	0.000000	0	NA
nodefactor.University_affiliation.BYU	-1.078431	3.670669	0	0.7689
nodefactor.University_affiliation.California State University at Stanislaus	-2.576070	1.948128	0	0.1861
nodefactor.University_affiliation.Cardiff University	-1.003211	2.360943	0	0.6709
nodefactor.University_affiliation.City University of Hong Kong	0.352172	1.716950	0	0.8375
nodefactor.University_affiliation.CMU	-0.945654	2.019097	0	0.6395
nodefactor.University_affiliation.College of Business Administration Kuwait	NA	0.000000	0	NA
nodefactor.University_affiliation.Colorado State University	NA	0.000000	0	NA
nodefactor.University_affiliation.Copenhagen Business School	NA	0.000000	0	NA
nodefactor.University_affiliation.Coriolis Labs	NA	0.000000	0	NA
nodefactor.University_affiliation.Curtin University of Technology	NA	0.000000	0	NA
nodefactor.University_affiliation.Deloitte MCS Ltd	NA	0.000000	0	NA
nodefactor.University_affiliation.Drexel University	-1.384960	3.562329	0	0.6974
nodefactor.University_affiliation.Emory University	-0.265748	1.833848	0	0.8848
nodefactor.University_affiliation.Erasmus University	NA	0.000000	0	NA
nodefactor.University_affiliation.ESAN University	NA	0.000000	0	NA
nodefactor.University_affiliation.ESC Rennes School of Business	0.787685	2.317379	0	0.7339
nodefactor.University_affiliation.ESSEC Business School	2.385786	3.275275	0	0.4664
nodefactor.University_affiliation.Florida State University	2.101594	1.700279	0	0.2165
nodefactor.University_affiliation.Fundação Getúlio Vargas	NA	0.000000	0	NA
nodefactor.University_affiliation.Georgia State University	1.477615	1.656367	0	0.3724
nodefactor.University_affiliation.HealthPartners Institute for Education and Research	-1.537871	5.290545	0	0.7713
nodefactor.University_affiliation.HEC Montreal	NA	0.000000	0	NA
nodefactor.University_affiliation.IMT Ghaziabad	NA	0.000000	0	NA
nodefactor.University_affiliation.Indiana University	1.519480	2.604164	0	0.5596
nodefactor.University_affiliation.INSEAD	NA	0.000000	0	NA
nodefactor.University_affiliation.Johns Hopkins University	-2.406786	2.551959	0	0.3456
nodefactor.University_affiliation.Keele University	0.501034	1.561879	0	0.7484
nodefactor.University_affiliation.Konyang University	NA	0.000000	0	NA
nodefactor.University_affiliation.Korea Advanced Institute of Science and Technology	NA	0.000000	0	NA
nodefactor.University_affiliation.London School of Economics	-1.100217	1.186567	0	0.3538
nodefactor.University_affiliation.Louisiana State University	NA	0.000000	0	NA
nodefactor.University_affiliation.Ludwig-Maximilian-University Munich	NA	0.000000	0	NA
nodefactor.University_affiliation.Miami University	NA	0.000000	0	NA
nodefactor.University_affiliation.Michigan State University	-2.416992	2.509006	0	0.3354
nodefactor.University_affiliation.Millsaps College	NA	0.000000	0	NA
nodefactor.University_affiliation.Monash University	NA	0.000000	0	NA
nodefactor.University_affiliation.NUS	1.342298	1.539186	0	0.3832
nodefactor.University_affiliation.NYU	0.519274	1.899783	0	0.7846

```
nodefactor.University_affiliation.Oklahama State                                   0.144726   3.597922      0  0.9679
nodefactor.University_affiliation.Pennsylvania State University                         NA   0.000000      0      NA
nodefactor.University_affiliation.Purdue University                                0.195188   3.642958      0  0.9573
nodefactor.University_affiliation.RMIT University                                       NA   0.000000      0      NA
nodefactor.University_affiliation.San Diego State University                      -1.359446   3.632611      0  0.7082
nodefactor.University_affiliation.Shells Chemicals Seraya Pvt Ltd                       NA   0.000000      0      NA
nodefactor.University_affiliation.Simon Fraser University                               NA   0.000000      0      NA
nodefactor.University_affiliation.Stevens Institute of Technology                       NA   0.000000      0      NA
nodefactor.University_affiliation.Tel-Aviv University                                   NA   0.000000      0      NA
nodefactor.University_affiliation.Tennessee Technological University                    NA   0.000000      0      NA
nodefactor.University_affiliation.Texas Christian University                       0.207945   3.377102      0  0.9509
nodefactor.University_affiliation.Texas Tech                                      -0.948205   2.012417      0  0.6375
nodefactor.University_affiliation.The Hong Kong Polytechnic University                  NA   0.000000      0      NA
nodefactor.University_affiliation.Three Gorges University                               NA   0.000000      0      NA
nodefactor.University_affiliation.Tsinghua University                                   NA   0.000000      0      NA
nodefactor.University_affiliation.Universidade do Vale do Rio dos Sinos                 NA   0.000000      0      NA
nodefactor.University_affiliation.Universidade NOVA ISEGI                               NA   0.000000      0      NA
nodefactor.University_affiliation.University of Amsterdam                               NA   0.000000      0      NA
nodefactor.University_affiliation.University of Antwerp                                 NA   0.000000      0      NA
nodefactor.University_affiliation.University of Arizona                            0.750638   2.247060      0  0.7383
nodefactor.University_affiliation.University of Calgary                            0.960474   1.803675      0  0.5944
nodefactor.University_affiliation.University of California Irvine                 -0.794825   2.191808      0  0.7169
nodefactor.University_affiliation.University of Cincinnati                        -0.056015   2.023896      0  0.9779
nodefactor.University_affiliation.University of Cologne                                 NA   0.000000      0      NA
nodefactor.University_affiliation.University of Florida                            0.763396   2.705339      0  0.7778
nodefactor.University_affiliation.University of Georgia                           -2.411889   2.530086      0  0.3405
nodefactor.University_affiliation.University of Hawaii Manoa                       0.714082   3.544430      0  0.8403
nodefactor.University_affiliation.University of Hong Kong                          2.053717   1.793042      0  0.2521
nodefactor.University_affiliation.University of Hull                                    NA   0.000000      0      NA
nodefactor.University_affiliation.University of Illinois                                NA   0.000000      0      NA
 [ reached getOption("max.print") -- omitted 35 rows ]
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 23340  on 16836  degrees of freedom
 Residual Deviance:  1487  on 16601  degrees of freedom

AIC: 1957    BIC: 3774    (Smaller is better.)
```

Figure A.2. Summary of main effects model with current Carnegie classification for author’s affiliated university

```
==========================
Summary of model fit
==========================

Formula:   n ~ edges + nodefactor("Gender") + nodefactor("National_affiliation") + 
    nodefactor("Dept_affiliation") + nodecov("Experience") + 
    nodefactor("PhD_granting_institute") + nodefactor("University_affiliation")

Iterations:  8 out of 20 

Monte Carlo MLE Results:
                                                                         Estimate Std. Error MCMC % p-value
edges                                                                   -5.247206   3.706313      0  0.1569
nodefactor.Gender.Male                                                   0.522408   0.619297      0  0.3989
nodefactor.National_affiliation.Non_american                            -0.744194   1.434565      0  0.6039
nodefactor.Dept_affiliation.Decision Sciences                           -0.168093   1.370860      0  0.9024
nodefactor.Dept_affiliation.Economics                                    2.256236   2.012959      0  0.2624
nodefactor.Dept_affiliation.Engineering Management                       0.129929   1.635872      0  0.9367
nodefactor.Dept_affiliation.Finance                                      0.129929   1.635872      0  0.9367
nodefactor.Dept_affiliation.IE                                           2.834314   2.204754      0  0.1986
nodefactor.Dept_affiliation.Industry                                     1.281018   2.227380      0  0.5652
nodefactor.Dept_affiliation.Innovation and Entrepreneurship              0.176984   2.178615      0  0.9353
nodefactor.Dept_affiliation.International Business                       0.631423   1.537220      0  0.6813
nodefactor.Dept_affiliation.IS                                           0.631423   0.581906      0  0.2779
nodefactor.Dept_affiliation.IT                                           0.383952   1.168220      0  0.7424
nodefactor.Dept_affiliation.IT Management                                0.589596   1.155438      0  0.6099
nodefactor.Dept_affiliation.Marketing                                    0.678478   1.569176      0  0.6655
nodefactor.Dept_affiliation.OM                                           0.506650   1.053063      0  0.6304
nodefactor.Dept_affiliation.Psychology                                  -0.802189   2.881132      0  0.7807
nodefactor.Dept_affiliation.Statistics                                   0.647108   1.545374      0  0.6754
nodefactor.Dept_affiliation.Technology Management                       -1.150984   1.847540      0  0.5333
nodecov.Experience                                                      -0.005228   0.021021      0  0.8036
nodefactor.PhD_granting_institute.Australian Graduate School of Management  0.214368   1.366412      0  0.8753
nodefactor.PhD_granting_institute.Cambridge University                   0.187441   1.655851      0  0.9099
nodefactor.PhD_granting_institute.Cardiff University                     0.104568   1.483638      0  0.9438
nodefactor.PhD_granting_institute.Case Western Reserve University       -1.049171   1.709830      0  0.5395
nodefactor.PhD_granting_institute.Chinese University of HongKong        -0.284928   2.194888      0  0.8967


nodefactor.PhD_granting_institute.City University of Hong Kong          -0.786422   2.244292      0  0.7260
nodefactor.PhD_granting_institute.Clemson University                    -0.592567   2.123189      0  0.7802
nodefactor.PhD_granting_institute.CMU                                   -2.620731   2.134937      0  0.2196
nodefactor.PhD_granting_institute.Columbia University                   -1.156806   2.590774      0  0.6552
nodefactor.PhD_granting_institute.Copenhagen Business School            -0.169769   1.340746      0  0.8992
nodefactor.PhD_granting_institute.Cranfield University                  -0.254023   1.698399      0  0.8811
nodefactor.PhD_granting_institute.Curtin University                     -0.501494   1.541800      0  0.7450
nodefactor.PhD_granting_institute.Ecole des Hautes Etudes en Sciences Sociales  1.121560   2.355558      0  0.6340
nodefactor.PhD_granting_institute.Ecole Superiere des Affaires          -0.538093   1.562161      0  0.7305
nodefactor.PhD_granting_institute.Erasmus University                    -0.517179   1.548848      0  0.7385
nodefactor.PhD_granting_institute.Florida Inernational University       -1.219546   2.179565      0  0.5758
nodefactor.PhD_granting_institute.Florida International University      -0.485809   1.537309      0  0.7520
nodefactor.PhD_granting_institute.Florida State University              -0.425259   2.049143      0  0.8356
nodefactor.PhD_granting_institute.Fundação Getúlio Vargas               -0.015685   1.424221      0  0.9912
nodefactor.PhD_granting_institute.Georgia State University              -0.532864   1.558418      0  0.7324
nodefactor.PhD_granting_institute.HKUST                                 -0.160227   2.283255      0  0.9441
nodefactor.PhD_granting_institute.HongKong Polytechnic University       -0.956007   2.287339      0  0.6760
nodefactor.PhD_granting_institute.Huazhong University of Science and Technology -2.162906   2.608047      0  0.4069
nodefactor.PhD_granting_institute.IIT India                             -0.446172   2.492498      0  0.8579
nodefactor.PhD_granting_institute.Indiana University                    -2.464588   2.313125      0  0.2867
nodefactor.PhD_granting_institute.Institute of France Telecom           -0.755051   2.237654      0  0.7358
nodefactor.PhD_granting_institute.Johns Hopkins University              -0.384492   1.972499      0  0.8455
nodefactor.PhD_granting_institute.KAIST Business School                 -0.833477   1.765415      0  0.6369
nodefactor.PhD_granting_institute.Katholieke University Leuven                 NA   0.000000      0      NA
nodefactor.PhD_granting_institute.Kent State University                 -2.592503   2.379640      0  0.2760
nodefactor.PhD_granting_institute.Kuwait University                      0.020914   1.425307      0  0.9883
nodefactor.PhD_granting_institute.LSE                                   -3.298798   3.455891      0  0.3398
nodefactor.PhD_granting_institute.McGill University                            NA   0.000000      0      NA
nodefactor.PhD_granting_institute.Medical College of Wisconsin          -0.631331   2.067680      0  0.7601
nodefactor.PhD_granting_institute.MIT                                   -1.695800   1.737486      0  0.3291
nodefactor.PhD_granting_institute.Monash University                     -0.501494   1.541800      0  0.7450
nodefactor.PhD_granting_institute.NA                                     0.840634   1.541402      0  0.5855
nodefactor.PhD_granting_institute.Northwestern University               -1.219546   2.602870      0  0.6394
nodefactor.PhD_granting_institute.Nova Information Management School    -0.517179   1.548848      0  0.7385
nodefactor.PhD_granting_institute.NUS                                   -1.219058   2.009325      0  0.5441
```

```
nodefactor.PhD_granting_institute.NYU                                          -0.597620  1.503371      0  0.6910
nodefactor.PhD_granting_institute.Pennsylvinia State University                -3.849191  2.210831      0  0.0817 .
nodefactor.PhD_granting_institute.Purdue University                            -0.558412  1.526307      0  0.7145
nodefactor.PhD_granting_institute.Queensland University of Technology           0.315817  2.099429      0  0.8804
nodefactor.PhD_granting_institute.RMIT                                         -0.480581  1.839849      0  0.7939
nodefactor.PhD_granting_institute.Stevens Institute of Technology               0.861247  2.203244      0  0.6959
nodefactor.PhD_granting_institute.Syracuse University                          -2.424276  1.905187      0  0.2032
nodefactor.PhD_granting_institute.Technion Israel Institute of Technology             NA  0.000000      0      NA
nodefactor.PhD_granting_institute.Texas Tech                                    0.214201  1.910098      0  0.9107
nodefactor.PhD_granting_institute.UBC                                          -2.797786  1.788241      0  0.1177
nodefactor.PhD_granting_institute.UC Berkeley                                  -0.749823  2.237237      0  0.7375
nodefactor.PhD_granting_institute.UC Irvine                                    -1.318885  2.260585      0  0.5596
nodefactor.PhD_granting_institute.UCLA                                         -0.464895  1.535335      0  0.7620
nodefactor.PhD_granting_institute.Universidade do Vale do Rio dos Sinos        -0.501494  1.367860      0  0.7139
nodefactor.PhD_granting_institute.Universita Cattolica del Sacro Cuore         -1.794238  2.544667      0  0.4808
nodefactor.PhD_granting_institute.University of Alberta                         0.879728  1.930641      0  0.6486
nodefactor.PhD_granting_institute.University of Arizona                        -1.245688  1.670273      0  0.4558
nodefactor.PhD_granting_institute.University of Bradford                       -0.548549  2.119148      0  0.7958
nodefactor.PhD_granting_institute.University of Cambridge                       0.564146  1.963089      0  0.7738
nodefactor.PhD_granting_institute.University of Cape Town                      -0.501494  1.541800      0  0.7450
nodefactor.PhD_granting_institute.University of Chicago                        -1.520245  1.801607      0  0.3988
nodefactor.PhD_granting_institute.University of Cincinnati                     -3.520028  1.953417      0  0.0716 .
nodefactor.PhD_granting_institute.University of Cologne                         1.824234  2.177144      0  0.4021
nodefactor.PhD_granting_institute.University of Georgia                        -2.652101  1.587245      0  0.0948 .
nodefactor.PhD_granting_institute.University of Houston                         0.436470  1.835735      0  0.8121
nodefactor.PhD_granting_institute.University of Iowa                           -3.118283  2.014019      0  0.1216
nodefactor.PhD_granting_institute.University of Kentucky                       -1.895283  2.983261      0  0.5252
nodefactor.PhD_granting_institute.University of Leeds                                 NA  0.000000      0      NA
nodefactor.PhD_granting_institute.University of Liechtenstein                  -0.590376  1.614233      0  0.7146
nodefactor.PhD_granting_institute.University of Liverpool                      -0.553778  1.575020      0  0.7251
nodefactor.PhD_granting_institute.University of Ljublijana                     -1.444980  2.442925      0  0.5542
nodefactor.PhD_granting_institute.University of Maryland                       -1.105230  2.281217      0  0.6280
nodefactor.PhD_granting_institute.University of Michigan                       -2.509361  1.851220      0  0.1753
nodefactor.PhD_granting_institute.University of Minnesota                       0.955610  1.711587      0  0.5766
nodefactor.PhD_granting_institute.University of Missouri                       -1.025938  2.052258      0  0.6171
nodefactor.PhD_granting_institute.University of Muenster                        0.558917  1.358451      0  0.6808
nodefactor.PhD_granting_institute.University of Munich                         -1.969456  2.472448      0  0.4257
nodefactor.PhD_granting_institute.University of Nottingham                     -1.895283  2.983261      0  0.5252
nodefactor.PhD_granting_institute.University of Oklahama                       -0.428754  2.112019      0  0.8391
```

```
nodefactor.PhD_granting_institute.University of Oulu                              -1.633423   2.183161      0  0.4544
nodefactor.PhD_granting_institute.University of Pennsylvinia                      -0.428464   1.620452      0  0.7915
nodefactor.PhD_granting_institute.University of Pittsburgh                        -1.015147   1.624988      0  0.5322
nodefactor.PhD_granting_institute.University of Rochester                         -0.590318   1.595286      0  0.7114
nodefactor.PhD_granting_institute.University of Sourthern California              -0.517179   1.548848      0  0.7385
nodefactor.PhD_granting_institute.University of South Carolina                    -1.245688   2.194138      0  0.5702
nodefactor.PhD_granting_institute.University of Strathclyde                        0.950778   3.041741      0  0.7546
nodefactor.PhD_granting_institute.University of Toronto                            0.214368   1.366412      0  0.8753
nodefactor.PhD_granting_institute.University of Utah                              -0.501494   1.541800      0  0.7450
nodefactor.PhD_granting_institute.University of Washington                        -1.689435   1.778308      0  0.3421
nodefactor.PhD_granting_institute.University of Wisconsin Milwaukee               -0.723280   2.013303      0  0.7194
nodefactor.PhD_granting_institute.University of Wollongong                         0.015685   1.424221      0  0.9912
nodefactor.PhD_granting_institute.Uppsala University                              -0.428297   1.542940      0  0.7813
nodefactor.PhD_granting_institute.UT Arlington                                    -1.775030   3.077858      0  0.5641
nodefactor.PhD_granting_institute.UT Austin                                       -0.386637   1.370820      0  0.7779
nodefactor.PhD_granting_institute.UT Dallas                                       -0.495389   2.082382      0  0.8120
nodefactor.PhD_granting_institute.Vrije University                                -0.321992   1.774816      0  0.8560
nodefactor.PhD_granting_institute.Washington State University                     -1.292743   2.232631      0  0.5626
nodefactor.PhD_granting_institute.Yonsei University                                0.621658   1.307871      0  0.6346
nodefactor.University_affiliation.AT&T Lab                                        -1.047467   1.896012      0  0.5806
nodefactor.University_affiliation.Auburn University                                      NA   0.000000      0      NA
nodefactor.University_affiliation.Australian National University                         NA   0.000000      0      NA
nodefactor.University_affiliation.Bentley University                               0.491037   1.538520      0  0.7496
nodefactor.University_affiliation.Blue Slate Solutions                             1.140632   2.519070      0  0.6507
nodefactor.University_affiliation.Brown University                                 1.211168   1.571666      0  0.4409
nodefactor.University_affiliation.Brunel University                                      NA   0.000000      0      NA
nodefactor.University_affiliation.BYU                                              0.330222   1.271205      0  0.7950
nodefactor.University_affiliation.California State University at Stanislaus       -0.624536   1.940104      0  0.7475
nodefactor.University_affiliation.Cardiff University                              -1.003062   2.351387      0  0.6697
nodefactor.University_affiliation.Chinese University of Hong Kong                        NA   0.000000      0      NA
nodefactor.University_affiliation.City University of Hong Kong                     0.284928   1.671258      0  0.8646
nodefactor.University_affiliation.CMU                                             -0.796311   1.705159      0  0.6405
nodefactor.University_affiliation.College of Business Administration Kuwait              NA   0.000000      0      NA
nodefactor.University_affiliation.Colorado State University                              NA   0.000000      0      NA
nodefactor.University_affiliation.Copenhagen Business School                             NA   0.000000      0      NA
nodefactor.University_affiliation.Coriolis Labs                                          NA   0.000000      0      NA

nodefactor.University_affiliation.Curtin University of Technology                        NA   0.000000      0      NA
nodefactor.University_affiliation.Deloitte MCS Ltd                                       NA   0.000000      0      NA
nodefactor.University_affiliation.Drexel University                               -0.041827   1.432728      0  0.9767
nodefactor.University_affiliation.Emory University                                -0.143190   1.568826      0  0.9273
nodefactor.University_affiliation.Erasmus University                                     NA   0.000000      0      NA
nodefactor.University_affiliation.ESAN University                                        NA   0.000000      0      NA
nodefactor.University_affiliation.ESC Rennes School of Business                    0.639138   2.233352      0  0.7747
nodefactor.University_affiliation.ESSEC Business School                            3.274500   2.429590      0  0.1778
nodefactor.University_affiliation.Florida State University                         1.907907   1.617936      0  0.2383
nodefactor.University_affiliation.Fundação Getúlio Vargas                                NA   0.000000      0      NA
nodefactor.University_affiliation.Georgia State University                         1.393870   1.599981      0  0.3837
nodefactor.University_affiliation.HealthPartners Institute for Education and Research -0.741544   2.690546      0  0.7829
nodefactor.University_affiliation.HEC Montreal                                           NA   0.000000      0      NA
nodefactor.University_affiliation.IMT Ghaziabad                                    0.160779   2.644335      0  0.9515
nodefactor.University_affiliation.Indiana University                               1.229356   2.495736      0  0.6223
nodefactor.University_affiliation.INSEAD                                                 NA   0.000000      0      NA
nodefactor.University_affiliation.Johns Hopkins University                        -2.128101   1.841518      0  0.2479
nodefactor.University_affiliation.Keele University                                 0.470124   1.535397      0  0.7595
nodefactor.University_affiliation.Konyang University                                     NA   0.000000      0      NA
nodefactor.University_affiliation.Korea Advanced Institute of Science and Technology     NA   0.000000      0      NA
nodefactor.University_affiliation.London School of Economics                      -1.076096   1.180564      0  0.3620
nodefactor.University_affiliation.Louisiana State University                             NA   0.000000      0      NA
nodefactor.University_affiliation.Ludwig-Maximilian-University Munich                    NA   0.000000      0      NA
nodefactor.University_affiliation.Miami University                                       NA   0.000000      0      NA
nodefactor.University_affiliation.Michigan State University                       -2.149014   1.832188      0  0.2408
nodefactor.University_affiliation.Millsaps College                                       NA   0.000000      0      NA
nodefactor.University_affiliation.Monash University                                      NA   0.000000      0      NA
nodefactor.University_affiliation.NUS                                              1.219058   1.418788      0  0.3902
nodefactor.University_affiliation.NYU                                              0.648200   1.647371      0  0.6940
nodefactor.University_affiliation.Oklahama State                                   1.208443   2.522966      0  0.6320
nodefactor.University_affiliation.Pennsylvania State University                          NA   0.000000      0      NA
nodefactor.University_affiliation.Purdue University                                1.354129   2.252102      0  0.5477
nodefactor.University_affiliation.RMIT University                                        NA   0.000000      0      NA
nodefactor.University_affiliation.San Diego State University                       0.010457   1.423446      0  0.9941
nodefactor.University_affiliation.Shells Chemicals Seraya Pvt Ltd                        NA   0.000000      0      NA
```

```
nodefactor.University_affiliation.Simon Fraser University                        NA  0.000000      0      NA
nodefactor.University_affiliation.Stevens Institute of Technology                NA  0.000000      0      NA
nodefactor.University_affiliation.Tel-Aviv University                            NA  0.000000      0      NA
nodefactor.University_affiliation.Tennessee Technological University             NA  0.000000      0      NA
nodefactor.University_affiliation.Texas Christian University               1.380271  1.714948      0  0.4209
nodefactor.University_affiliation.Texas Tech                              -0.801539  1.705566      0  0.6384
nodefactor.University_affiliation.The Hong Kong Polytechnic University           NA  0.000000      0      NA
nodefactor.University_affiliation.Three Gorges University                        NA  0.000000      0      NA
nodefactor.University_affiliation.Tsinghua University                            NA  0.000000      0      NA
nodefactor.University_affiliation.Universidade do Vale do Rio dos Sinos          NA  0.000000      0      NA
nodefactor.University_affiliation.Universidade NOVA ISEGI                        NA  0.000000      0      NA
nodefactor.University_affiliation.University of Amsterdam                        NA  0.000000      0      NA
nodefactor.University_affiliation.University of Antwerp                          NA  0.000000      0      NA
nodefactor.University_affiliation.University of Arizona                    1.083486  1.152330      0  0.3471
nodefactor.University_affiliation.University of Calgary                    1.054306  1.760935      0  0.5494
nodefactor.University_affiliation.University of California Irvine        -0.682048  1.974587      0  0.7298
nodefactor.University_affiliation.University of Cincinnati                 0.081405  1.759366      0  0.9631
nodefactor.University_affiliation.University of Cologne                          NA  0.000000      0      NA
nodefactor.University_affiliation.University of Florida                    1.258444  1.803001      0  0.4852
nodefactor.University_affiliation.University of Georgia                   -2.138557  1.836378      0  0.2442
nodefactor.University_affiliation.University of Hawaii Manoa               1.860851  2.180949      0  0.3935
nodefactor.University_affiliation.University of Hong Kong                  1.830633  1.085023      0  0.0916 .
nodefactor.University_affiliation.University of Hull                             NA  0.000000      0      NA
nodefactor.University_affiliation.University of Illinois                   1.882228  2.179660      0  0.3879
nodefactor.University_affiliation.University of Iowa                       2.535534  1.398532      0  0.0699 .
nodefactor.University_affiliation.University of Liechtenstein                    NA  0.000000      0      NA
nodefactor.University_affiliation.University of Ljubljana                        NA  0.000000      0      NA
nodefactor.University_affiliation.University of Louisville                -0.107635  1.681239      0  0.9490
nodefactor.University_affiliation.University of Maryland                   1.609692  1.637500      0  0.3256
nodefactor.University_affiliation.University of Massachusetts Boston      -0.634992  1.944375      0  0.7440
 [ reached getOption("max.print") -- omitted 29 rows ]
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 23340  on 16836  degrees of freedom
 Residual Deviance:  1487  on 16607  degrees of freedom

AIC: 1945    BIC: 3716    (Smaller is better.)
```

Figure A.3 Summary of main effects model without current Carnegie classification for author's affiliated university

```
==========================
Summary of model fit
==========================

Formula:   n ~ edges + nodematch("Gender") + nodematch("National_affiliation") +
    nodematch("Dept_affiliation") + nodematch("Experience") +
    nodematch("Carnegie_classification") + nodematch("PhD_granting_institute") +
    nodematch("University_affiliation")

Iterations:  7 out of 20

Monte Carlo MLE Results:
                                  Estimate Std. Error MCMC % p-value
edges                              -5.7920     0.2272      0  <1e-04 ***
nodematch.Gender                    0.1580     0.1903      0  0.4066
nodematch.National_affiliation      0.1634     0.2377      0  0.4918
nodematch.Dept_affiliation          0.7237     0.1825      0  <1e-04 ***
nodematch.Experience               -0.1879     0.4719      0  0.6904
nodematch.Carnegie_classification   0.3935     0.2254      0  0.0808 .
nodematch.PhD_granting_institute    1.9097     0.2909      0  <1e-04 ***
nodematch.University_affiliation    2.6381     0.2755      0  <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 23340  on 16836  degrees of freedom
 Residual Deviance:  1431  on 16828  degrees of freedom

AIC: 1447    BIC: 1508    (Smaller is better.)
```

Figure A.4. Summary of homophily model with current Carnegie classification for author's affiliated university

```
==========================
Summary of model fit
==========================

Formula:   n ~ edges + nodematch("Gender") + nodematch("National_affiliation") +
    nodematch("Dept_affiliation") + nodematch("Experience") +
    nodematch("PhD_granting_institute") + nodematch("University_affiliation")

Iterations:  7 out of 20

Monte Carlo MLE Results:
                                 Estimate Std. Error MCMC % p-value
edges                             -5.7616     0.2277      0  <1e-04 ***
nodematch.Gender                   0.1652     0.1902      0  0.3850
nodematch.National_affiliation     0.3880     0.1963      0  0.0481 *
nodematch.Dept_affiliation         0.7266     0.1823      0  <1e-04 ***
nodematch.Experience              -0.1518     0.4717      0  0.7476
nodematch.PhD_granting_institute   1.9061     0.2901      0  <1e-04 ***
nodematch.University_affiliation   2.7452     0.2695      0  <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 23340  on 16836  degrees of freedom
 Residual Deviance:  1434  on 16829  degrees of freedom

AIC: 1448    BIC: 1502    (Smaller is better.)
```

Figure A.5. Summary of homophily model without current Carnegie classification for author's affiliated university

```
==========================
Summary of model fit
==========================

Formula:   n ~ edges + nodematch("Gender", diff = T) + nodematch("National_affiliation",
    diff = T) + nodematch("Dept_affiliation", diff = T) + nodecov("Experience") +
    nodematch("PhD_granting_institute", diff = T) + nodematch("University_affiliation",
    diff = T) + nodematch("Carnegie_classification", diff = T)

Iterations:  15 out of 20

Monte Carlo MLE Results:
                                                                              Estimate Std. Error MCMC %  p-value
edges                                                                        -5.619e+00  3.513e-01      0  < 1e-04 ***
nodematch.Gender.Female                                                       1.528e-01  4.499e-01      0 0.734088
nodematch.Gender.Male                                                         1.032e-01  2.080e-01      0 0.619635
nodematch.National_affiliation.American                                       6.713e-02  2.826e-01      0 0.812266
nodematch.National_affiliation.Non_american                                   9.667e-02  5.310e-01      0 0.855541
nodematch.Dept_affiliation.CSE                                                     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.Decision Sciences                                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.Economics                                         -1.109e+01  4.458e+02      0 0.980151
nodematch.Dept_affiliation.Engineering Management                                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.Finance                                           -1.131e+01  1.692e+03      0 0.994668
nodematch.Dept_affiliation.IE                                                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.Industry                                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.Innovation and Entrepreneurship                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.International Business                                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.IS                                                 6.163e-01  2.053e-01      0 0.002692 **
nodematch.Dept_affiliation.IT                                                 1.575e+00  1.194e+00      0 0.187351
nodematch.Dept_affiliation.IT Management                                      4.816e+00  1.249e+00      0 0.000115 ***
nodematch.Dept_affiliation.Marketing                                          3.763e+00  5.751e-01      0  < 1e-04 ***
nodematch.Dept_affiliation.OM                                                 2.816e+00  7.869e-01      0 0.000347 ***
nodematch.Dept_affiliation.Psychology                                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.Statistics                                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.Technology Management                                   -Inf  0.000e+00      0  < 1e-04 ***
nodecov.Experience                                                           -4.529e-03  7.109e-03      0 0.524055
nodematch.PhD_granting_institute.Athens University of Economics and Business       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Australian Graduate School of Management          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Cambridge University                              -Inf  0.000e+00      0  < 1e-04 ***

nodematch.PhD_granting_institute.Cardiff University                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Case Western Reserve University                   -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Chinese University of HongKong                    -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.City University of Hong Kong                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Clemson University                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.CMU                                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Columbia University                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Copenhagen Business School                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Cranfield University                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Curtin University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Ecole des Hautes Etudes en Sciences Sociales      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Ecole Superiere des Affaires                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Erasmus University                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Florida Inernational University                   -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Florida International University                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Florida State University                     2.229e+01  2.400e+03      0 0.992590
nodematch.PhD_granting_institute.Fundação Getúlio Vargas                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Georgia State University                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.HKUST                                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.HongKong Polytechnic University                   -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Huazhong University of Science and Technology     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.IIT India                                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Indiana University                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Institute of France Telecom                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Johns Hopkins University                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.KAIST Business School                        2.152e+01  2.400e+03      0 0.992843
nodematch.PhD_granting_institute.Katholieke University Leuven                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Kent State University                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Kuwait University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.LSE                                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.McGill University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Medical College of Wisconsin                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.MIT                                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Monash University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.NA                                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Northwestern University                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Nova Information Management School                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.NUS                                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.NYU                                               -Inf  0.000e+00      0  < 1e-04 ***
```

```
nodematch.PhD_granting_institute.Pennsylvinia State University                    -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Purdue University                           2.256e+00  1.170e+00      0 0.053794 .
nodematch.PhD_granting_institute.Queensland University of Technology              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.RMIT                                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Stevens Institute of Technology                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Syracuse University                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Technion Israel Institute of Technology          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Texas Tech                                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.UBC                                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.UC Berkeley                                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.UC Irvine                                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.UCLA                                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Universidade do Vale do Rio dos Sinos            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Universita Cattolica del Sacro Cuore             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Alberta                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Arizona                       2.268e+00  5.624e-01      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Bradford                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Cambridge                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Cape Town                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Chicago                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Cincinnati                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Cologne                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Georgia                       3.046e+00  9.568e-01      0 0.001457 **
nodematch.PhD_granting_institute.University of Houston                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Iowa                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Kentucky                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Leeds                         3.316e+01  2.936e+03      0 0.990991
nodematch.PhD_granting_institute.University of Liechtenstein                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Liverpool                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Ljublijana                    3.284e+01  2.441e+03      0 0.989264
nodematch.PhD_granting_institute.University of Maryland                      1.938e+01  2.400e+03      0 0.993557
nodematch.PhD_granting_institute.University of Michigan                      3.273e+00  8.256e-01      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Minnesota                     2.235e+00  8.725e-01      0 0.010439 *
nodematch.PhD_granting_institute.University of Missouri                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Muenster                      2.110e+01  2.400e+03      0 0.992983
nodematch.PhD_granting_institute.University of Munich                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Nottingham                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Oklahama                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Oulu                               -Inf  0.000e+00      0  < 1e-04 ***

nodematch.PhD_granting_institute.University of Pennsylvinia                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Pittsburgh                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Rochester                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Sourthern California               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of South Carolina                     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Strathclyde                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Toronto                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Utah                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Washington                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Wisconsin Milwaukee                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Wollongong                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Uppsala University                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.UT Arlington                                     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.UT Austin                                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.UT Dallas                                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Vrije University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Washington State University                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Yonsei University                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Arizona State University                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.AT&T Lab                                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Auburn University                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Australian National University                   -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Bentley University                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Blue Slate Solutions                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Brown University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Brunel University                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.BYU                                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.California State University at Stanislaus        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Cardiff University                          2.176e+01  1.660e+03      0 0.989542
nodematch.University_affiliation.Chinese University of Hong Kong                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.City University of Hong Kong                2.638e+00  7.880e-01      0 0.000815 ***
nodematch.University_affiliation.CMU                                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.College of Business Administration Kuwait   2.124e+01  2.400e+03      0 0.992936
nodematch.University_affiliation.Colorado State University                   2.098e+01  2.400e+03      0 0.993023
nodematch.University_affiliation.Copenhagen Business School                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Coriolis Labs                                    -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Curtin University of Technology                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Deloitte MCS Ltd                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Drexel University                                -Inf  0.000e+00      0  < 1e-04 ***
```

```
nodematch.University_affiliation.Emory University                                   -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Erasmus University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.ESAN University                                    -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.ESC Rennes School of Business                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.ESSEC Business School                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Florida State University                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Fundação Getúlio Vargas                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Georgia State University                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.HealthPartners Institute for Education and Research -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.HEC Montreal                                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.IMT Ghaziabad                                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Indiana University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.INSEAD                                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Johns Hopkins University                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Keele University                                   -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Konyang University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Korea Advanced Institute of Science and Technology -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.London School of Economics                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Louisiana State University                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Ludwig-Maximilian-University Munich                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Miami University                                   -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Michigan State University                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Millsaps College                                   -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Monash University                                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.NUS                                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.NYU                                           3.097e+00  6.691e-01      0  < 1e-04 ***
nodematch.University_affiliation.Oklahama State                                     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Pennsylvania State University                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Purdue University                                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.RMIT University                                    -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.San Diego State University                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Shells Chemicals Seraya Pvt Ltd                    -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Simon Fraser University                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Stevens Institute of Technology                    -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Tel-Aviv University                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Tennessee Technological University                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Texas Christian University                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Texas Tech                                    2.102e+01  2.400e+03      0 0.993012

nodematch.University_affiliation.The Hong Kong Polytechnic University               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Three Gorges University                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Tsinghua University                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Universidade do Vale do Rio dos Sinos              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Universidade NOVA ISEGI                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Amsterdam                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Antwerp                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Arizona                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Calgary                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of California Irvine                    -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Cincinnati                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Cologne                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Florida                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Georgia                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Hawaii Manoa                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Hong Kong                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Hull                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Illinois                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Iowa                            3.998e+00  8.521e-01      0  < 1e-04 ***
nodematch.University_affiliation.University of Liechtenstein                   3.867e+00  1.256e+00      0 0.002073 **
nodematch.University_affiliation.University of Ljubljana                              NA  0.000e+00      0       NA
nodematch.University_affiliation.University of Louisville                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Maryland                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Massachusetts Boston                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Massachusetts Dartmouth              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Massachusetts Lowell                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Michigan                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Minnesota                       3.796e+00  7.695e-01      0  < 1e-04 ***
nodematch.University_affiliation.University of Missouri                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of North Texas                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Pennsylvania                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Sheffield                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of South Carolina                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of South Florida                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Sussex                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Technology                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Texas Austin                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Texas Dallas                    2.047e+00  8.107e-01      0 0.011582 *
nodematch.University_affiliation.University of Texas El Paso                        -Inf  0.000e+00      0  < 1e-04 ***
```

```
nodematch.University_affiliation.University of Texas San Antonio                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Texas Southwestern Medical Center -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Toronto                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Utah                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Virginia                      3.557e+00  8.098e-01      0  < 1e-04 ***
nodematch.University_affiliation.University of Warwick                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Washington                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Wisconsin Milwaukee               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Wollongong                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Zurich                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Virginia Polytech                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Washington State University                     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Yonsei University                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.York University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Carnegie_classification.15                                         5.465e-01  3.018e-01      0 0.070215 .
nodematch.Carnegie_classification.16                                         9.256e-01  5.922e-01      0 0.118116
nodematch.Carnegie_classification.17                                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Carnegie_classification.18                                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Carnegie_classification.21                                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Carnegie_classification.25                                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Carnegie_classification.99                                         3.504e-01  5.114e-01      0 0.493273
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 23340  on 16836  degrees of freedom
 Residual Deviance:  1586  on 16595  degrees of freedom

AIC: 2068    BIC: 3931    (Smaller is better.)
```

Figure A.6. Summary of differential homophily model with current Carnegie classification for author’s affiliated university

```
==========================
Summary of model fit
==========================

Formula:   n ~ edges + nodematch("Gender", diff = T) + nodematch("National_affiliation",
    diff = T) + nodematch("Dept_affiliation", diff = T) + nodecov("Experience") +
    nodematch("PhD_granting_institute", diff = T) + nodematch("University_affiliation",
    diff = T)

Iterations:  15 out of 20

Monte Carlo MLE Results:
                                                                                     Estimate Std. Error MCMC %  p-value
edges                                                                               -5.636e+00  3.434e-01      0  < 1e-04 ***
nodematch.Gender.Female                                                              1.316e-01  4.508e-01      0 0.770265
nodematch.Gender.Male                                                                1.173e-01  2.075e-01      0 0.572041
nodematch.National_affiliation.American                                              3.498e-01  2.231e-01      0 0.116931
nodematch.National_affiliation.Non_american                                          4.032e-01  2.856e-01      0 0.158024
nodematch.Dept_affiliation.CSE                                                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.Decision Sciences                                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.Economics                                                -1.106e+01  4.429e+02      0 0.980071
nodematch.Dept_affiliation.Engineering Management                                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.Finance                                                  -1.129e+01  1.696e+03      0 0.994691
nodematch.Dept_affiliation.IE                                                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.Industry                                                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.Innovation and Entrepreneurship                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.International Business                                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.IS                                                        6.404e-01  2.019e-01      0 0.001516 **
nodematch.Dept_affiliation.IT                                                        1.569e+00  1.196e+00      0 0.189547
nodematch.Dept_affiliation.IT Management                                             4.805e+00  1.246e+00      0 0.000116 ***
nodematch.Dept_affiliation.Marketing                                                 3.772e+00  5.739e-01      0  < 1e-04 ***
nodematch.Dept_affiliation.OM                                                        2.948e+00  7.578e-01      0 0.000100 ***
nodematch.Dept_affiliation.Psychology                                                     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.Statistics                                                     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.Dept_affiliation.Technology Management                                          -Inf  0.000e+00      0  < 1e-04 ***
nodecov.Experience                                                                  -3.643e-03  6.991e-03      0 0.602329
nodematch.PhD_granting_institute.Athens University of Economics and Business              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Australian Graduate School of Management                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Cambridge University                                     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Cardiff University                                       -Inf  0.000e+00      0  < 1e-04 ***

nodematch.PhD_granting_institute.Case Western Reserve University                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Chinese University of HongKong                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.City University of Hong Kong                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Clemson University                                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.CMU                                                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Columbia University                                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Copenhagen Business School                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Cranfield University                                     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Curtin University                                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Ecole des Hautes Etudes en Sciences Sociales             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Ecole Superiere des Affaires                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Erasmus University                                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Florida Inernational University                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Florida International University                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Florida State University                            2.195e+01  2.400e+03      0 0.992700
nodematch.PhD_granting_institute.Fundação Getúlio Vargas                                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Georgia State University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.HKUST                                                    -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.HongKong Polytechnic University                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Huazhong University of Science and Technology            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.IIT India                                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Indiana University                                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Institute of France Telecom                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Johns Hopkins University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.KAIST Business School                               2.149e+01  2.400e+03      0 0.992854
nodematch.PhD_granting_institute.Katholieke University Leuven                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Kent State University                                    -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Kuwait University                                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.LSE                                                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.McGill University                                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Medical College of Wisconsin                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.MIT                                                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Monash University                                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.NA                                                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Northwestern University                                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Nova Information Management School                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.NUS                                                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.NYU                                                      -Inf  0.000e+00      0  < 1e-04 ***
```

```
nodematch.PhD_granting_institute.Pennsylvinia State University                     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Purdue University                             2.115e+00  1.149e+00      0 0.065584 .
nodematch.PhD_granting_institute.Queensland University of Technology                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.RMIT                                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Stevens Institute of Technology                    -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Syracuse University                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Technion Israel Institute of Technology            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Texas Tech                                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.UBC                                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.UC Berkeley                                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.UC Irvine                                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.UCLA                                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Universidade do Vale do Rio dos Sinos              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Universita Cattolica del Sacro Cuore               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Alberta                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Arizona                         2.225e+00  5.468e-01      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Bradford                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Cambridge                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Cape Town                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Chicago                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Cincinnati                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Cologne                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Georgia                         3.017e+00  9.428e-01      0 0.001379 **
nodematch.PhD_granting_institute.University of Houston                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Iowa                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Kentucky                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Leeds                           3.317e+01  2.939e+03      0 0.990993
nodematch.PhD_granting_institute.University of Liechtenstein                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Liverpool                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Ljubljana                       3.284e+01  2.440e+03      0 0.989263
nodematch.PhD_granting_institute.University of Maryland                        1.893e+01  2.400e+03      0 0.993706
nodematch.PhD_granting_institute.University of Michigan                        3.493e+00  8.155e-01      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Minnesota                       2.211e+00  8.649e-01      0 0.010600 *
nodematch.PhD_granting_institute.University of Missouri                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Muenster                        2.111e+01  2.400e+03      0 0.992981
nodematch.PhD_granting_institute.University of Munich                               -Inf  0.000e+00      0  < 1e-04 ***

nodematch.PhD_granting_institute.University of Nottingham                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Oklahama                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Oulu                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Pennsylvinia                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Pittsburgh                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Rochester                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Sourthern California                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of South Carolina                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Strathclyde                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Toronto                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Utah                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Washington                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Wisconsin Milwaukee                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.University of Wollongong                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Uppsala University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.UT Arlington                                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.UT Austin                                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.UT Dallas                                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Vrije University                                   -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Washington State University                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.PhD_granting_institute.Yonsei University                                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Arizona State University                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.AT&T Lab                                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Auburn University                                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Australian National University                     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Bentley University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Blue Slate Solutions                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Brown University                                   -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Brunel University                                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.BYU                                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.California State University at Stanislaus          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Cardiff University                            2.178e+01  1.643e+03      0 0.989419
nodematch.University_affiliation.Chinese University of Hong Kong                    -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.City University of Hong Kong                  2.641e+00  7.879e-01      0 0.000805 ***
nodematch.University_affiliation.CMU                                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.College of Business Administration Kuwait     2.126e+01  2.400e+03      0 0.992932
```

```
nodematch.University_affiliation.Colorado State University                  2.120e+01  2.400e+03      0 0.992951
nodematch.University_affiliation.Copenhagen Business School                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Coriolis Labs                                   -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Curtin University of Technology                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Deloitte MCS Ltd                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Drexel University                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Emory University                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Erasmus University                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.ESAN University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.ESC Rennes School of Business                   -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.ESSEC Business School                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Florida State University                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Fundação Getúlio Vargas                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Georgia State University                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.HealthPartners Institute for Education and Research  -Inf  0.000e+00  0  < 1e-04 ***
nodematch.University_affiliation.HEC Montreal                                    -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.IMT Ghaziabad                                   -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Indiana University                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.INSEAD                                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Johns Hopkins University                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Keele University                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Konyang University                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Korea Advanced Institute of Science and Technology  -Inf  0.000e+00   0  < 1e-04 ***
nodematch.University_affiliation.London School of Economics                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Louisiana State University                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Ludwig-Maximilian-University Munich             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Miami University                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Michigan State University                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Millsaps College                                -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Monash University                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.NUS                                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.NYU                                        3.327e+00  6.549e-01      0  < 1e-04 ***
nodematch.University_affiliation.Oklahama State                                  -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Pennsylvania State University                   -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Purdue University                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.RMIT University                                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.San Diego State University                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Shells Chemicals Seraya Pvt Ltd                 -Inf  0.000e+00      0  < 1e-04 ***

nodematch.University_affiliation.Shells Chemicals Seraya Pvt Ltd                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Simon Fraser University                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Stevens Institute of Technology                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Tel-Aviv University                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Tennessee Technological University              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Texas Christian University                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Texas Tech                                 2.123e+01  2.400e+03      0 0.992942
nodematch.University_affiliation.The Hong Kong Polytechnic University            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Three Gorges University                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Tsinghua University                             -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Universidade do Vale do Rio dos Sinos           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Universidade NOVA ISEGI                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Amsterdam                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Antwerp                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Arizona                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Calgary                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of California Irvine                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Cincinnati                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Cologne                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Florida                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Georgia                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Hawaii Manoa                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Hong Kong                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Hull                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Illinois                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Iowa                         3.976e+00  8.481e-01      0  < 1e-04 ***
nodematch.University_affiliation.University of Liechtenstein                3.869e+00  1.255e+00      0 0.002056 **
nodematch.University_affiliation.University of Ljubljana                           NA  0.000e+00      0       NA
nodematch.University_affiliation.University of Louisville                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Maryland                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Massachusetts Boston              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Massachusetts Dartmouth           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Massachusetts Lowell              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Michigan                          -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Minnesota                    4.018e+00  7.573e-01      0  < 1e-04 ***
nodematch.University_affiliation.University of Missouri                          -Inf  0.000e+00      0  < 1e-04 ***
```

```
nodematch.University_affiliation.University of North Texas                       -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Pennsylvania                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Sheffield                         -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of South Carolina                    -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of South Florida                     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Sussex                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Technology                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Texas Austin                      -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Texas Dallas                 2.275e+00  7.956e-01      0 0.004242 **
nodematch.University_affiliation.University of Texas El Paso                     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Texas San Antonio                 -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Texas Southwestern Medical Center -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Toronto                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Utah                              -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Virginia                     3.788e+00  7.968e-01      0  < 1e-04 ***
nodematch.University_affiliation.University of Warwick                           -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Washington                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Wisconsin Milwaukee               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Wollongong                        -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.University of Zurich                            -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Virginia Polytech                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Washington State University                     -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.Yonsei University                               -Inf  0.000e+00      0  < 1e-04 ***
nodematch.University_affiliation.York University                                 -Inf  0.000e+00      0  < 1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 23340  on 16836  degrees of freedom
 Residual Deviance:  1582  on 16602  degrees of freedom

AIC: 2050    BIC: 3859    (Smaller is better.)
```

Figure A.7. Summary of differential homophily model without current Carnegie classification for author's affiliated university

```
Call:
glm(formula = poisson_data_sna$degree ~ poisson_data_sna$gender + 
    poisson_data_sna$continental_affiliation + poisson_data_sna$university_affiliation + 
    poisson_data_sna$dept_affiliation + poisson_data_sna$experience + 
    poisson_data_sna$current_carnegie + poisson_data_sna$phd_granting_inst, 
    family = poisson(link = "log"), data = poisson_data_sna)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-0.8762   0.0000   0.0000   0.0000   0.7402  

Coefficients: (52 not defined because of singularities)
                                                                                    Estimate Std. Error z value Pr(>|z|)  
(Intercept)                                                                         -2.29359    1.59545  -1.438   0.1506  
poisson_data_sna$genderMale                                                         -0.51676    1.13467  -0.455   0.6488  
poisson_data_sna$continental_affiliationNon_american                                 2.88104    2.12685   1.355   0.1755  
poisson_data_sna$university_affiliationAT&T Lab                                      3.26458    2.31438   1.411   0.1584  
poisson_data_sna$university_affiliationAuburn University                             0.25372    3.39389   0.075   0.9404  
poisson_data_sna$university_affiliationAustralian National University                0.56150    1.66933   0.336   0.7366  
poisson_data_sna$university_affiliationBentley University                           -0.59575    1.78739  -0.333   0.7389  
poisson_data_sna$university_affiliationBlue Slate Solutions                          1.56582    2.13384   0.734   0.4631  
poisson_data_sna$university_affiliationBrown University                              1.86588    1.90527   0.979   0.3274  
poisson_data_sna$university_affiliationBrunel University                            -2.25018    1.83484  -1.226   0.2201  
poisson_data_sna$university_affiliationBYU                                          -0.55679    3.22662  -0.173   0.8630  
poisson_data_sna$university_affiliationCalifornia State University at Stanislaus     2.80136    2.35422   1.190   0.2341  
poisson_data_sna$university_affiliationCardiff University                           -1.42054    2.14451  -0.662   0.5077  
poisson_data_sna$university_affiliationChinese University of Hong Kong               1.04511    1.49475   0.699   0.4844  
poisson_data_sna$university_affiliationCity University of Hong Kong                 -0.73595    0.91533  -0.804   0.4214  
poisson_data_sna$university_affiliationCMU                                           1.03141    1.64219   0.628   0.5300  
poisson_data_sna$university_affiliationCollege of Business Administration Kuwait    -1.23624    1.65941  -0.745   0.4563  
poisson_data_sna$university_affiliationColorado State University                     1.48019    1.57553   0.939   0.3475  
poisson_data_sna$university_affiliationCopenhagen Business School                   -0.99435    1.40863  -0.706   0.4803  
poisson_data_sna$university_affiliationCoriolis Labs                                 4.77762    2.70098   1.769   0.0769  
poisson_data_sna$university_affiliationCurtin University of Technology               0.37914    0.91367   0.415   0.6782  
poisson_data_sna$university_affiliationDeloitte MCS Ltd                              4.77762    2.70098   1.769   0.0769  
poisson_data_sna$university_affiliationDrexel University                            -0.74920    3.09592  -0.242   0.8088  
poisson_data_sna$university_affiliationEmory University                              1.59292    1.49256   1.067   0.2859  
poisson_data_sna$university_affiliationErasmus University                           -0.75897    1.22848  -0.618   0.5367  
poisson_data_sna$university_affiliationESAN University                               0.41863    0.91307   0.458   0.6466  

poisson_data_sna$university_affiliationESC Rennes School of Business                 0.34562    2.13984   0.162   0.8717  
poisson_data_sna$university_affiliationESSEC Business School                         0.01009    3.07891   0.003   0.9974  
poisson_data_sna$university_affiliationFlorida State University                      2.12099    1.74940   1.212   0.2254  
poisson_data_sna$university_affiliationFundação Getúlio Vargas                      -1.32839    1.63043  -0.815   0.4152  
poisson_data_sna$university_affiliationGeorgia State University                      0.95730    1.61970   0.591   0.5545  
poisson_data_sna$university_affiliationHealthPartners Institute for Education and Research 3.84717 2.37312 1.621 0.1050  
poisson_data_sna$university_affiliationHEC Montreal                                  0.33089    1.86027   0.178   0.8588  
poisson_data_sna$university_affiliationIMT Ghaziabad                                -2.32917    1.84142  -1.265   0.2059  
poisson_data_sna$university_affiliationIndiana University                            2.51137    2.67060   0.940   0.3470  
poisson_data_sna$university_affiliationINSEAD                                       -0.83796    1.88265  -0.445   0.6563  
poisson_data_sna$university_affiliationJohns Hopkins University                     -1.13398    2.02128  -0.561   0.5748  
poisson_data_sna$university_affiliationKeele University                             -1.38436    1.98729  -0.697   0.4860  
poisson_data_sna$university_affiliationKonyang University                            0.37914    0.91367   0.415   0.6782  
poisson_data_sna$university_affiliationKorea Advanced Institute of Science and Technology -0.62733 1.22848 -0.511 0.6096  
poisson_data_sna$university_affiliationLondon School of Economics                   -1.92340    1.69411  -1.135   0.2562  
poisson_data_sna$university_affiliationLouisiana State University                    2.39713    1.42944   1.677   0.0935 .
poisson_data_sna$university_affiliationLudwig-Maximilian-University Munich          -2.49263    2.52227  -0.988   0.3230  
poisson_data_sna$university_affiliationMiami University                              0.77048    3.17192   0.243   0.8081  
poisson_data_sna$university_affiliationMichigan State University                     1.54601    1.57424   0.982   0.3261  
poisson_data_sna$university_affiliationMillsaps College                              1.29588    1.60973   0.805   0.4208  
poisson_data_sna$university_affiliationMonash University                             0.37914    0.91367   0.415   0.6782  
poisson_data_sna$university_affiliationNUS                                          -0.02633    1.00073  -0.026   0.9790  
poisson_data_sna$university_affiliationNYU                                           1.89490    1.31488   1.441   0.1496  
poisson_data_sna$university_affiliationOklahama State                                1.81482    3.29561   0.551   0.5819  
poisson_data_sna$university_affiliationPennsylvania State University                 1.48019    2.11714   0.699   0.4845  
poisson_data_sna$university_affiliationPurdue University                             1.91488    3.09704   0.618   0.5364  
poisson_data_sna$university_affiliationRMIT University                              -0.71948    1.22534  -0.587   0.5571  
poisson_data_sna$university_affiliationSan Diego State University                   -0.61756    3.09696  -0.199   0.8419  
poisson_data_sna$university_affiliationShells Chemicals Seraya Pvt Ltd               2.98800    2.72069   1.098   0.2721  
poisson_data_sna$university_affiliationSimon Fraser University                      -0.75897    1.22848  -0.618   0.5367  
poisson_data_sna$university_affiliationStevens Institute of Technology               2.51746    3.51446   0.716   0.4738  
poisson_data_sna$university_affiliationTel-Aviv University                          -0.37682    2.43953  -0.154   0.8772  
poisson_data_sna$university_affiliationTennessee Technological University            3.08680    1.76009   1.754   0.0795 .
poisson_data_sna$university_affiliationTexas Christian University                    1.98070    2.73562   0.724   0.4690  
poisson_data_sna$university_affiliationTexas Tech                                    1.01825    1.64213   0.620   0.5352  
poisson_data_sna$university_affiliationThe Hong Kong Polytechnic University         -0.79846    1.23429  -0.647   0.5177  
```

```
poisson_data_sna$university_affiliationThe University of Hong Kong                -1.49335   1.21656  -1.228   0.2196
poisson_data_sna$university_affiliationThree Gorges University                    -2.28656   2.28161  -1.002   0.3163
poisson_data_sna$university_affiliationTsinghua University                        -2.85598   3.05017  -0.936   0.3491
poisson_data_sna$university_affiliationUniversidade do Vale do Rio dos Sinos      -0.31401   0.91367  -0.344   0.7311
poisson_data_sna$university_affiliationUniversidade NOVA ISEGI                     0.90030   1.46004   0.617   0.5375
poisson_data_sna$university_affiliationUniversity of Amsterdam                    -1.32271   1.66905  -0.792   0.4281
poisson_data_sna$university_affiliationUniversity of Antwerp                       0.90030   1.46004   0.617   0.5375
poisson_data_sna$university_affiliationUniversity of Arizona                       1.71417   1.48661   1.153   0.2489
poisson_data_sna$university_affiliationUniversity of Calgary                       0.22116   0.95143   0.232   0.8162
poisson_data_sna$university_affiliationUniversity of California Irvine             1.42753   1.58075   0.903   0.3665
poisson_data_sna$university_affiliationUniversity of Cincinnati                    1.58158   1.58727   0.996   0.3190
poisson_data_sna$university_affiliationUniversity of Cologne                      -1.56058   3.20960  -0.486   0.6268
poisson_data_sna$university_affiliationUniversity of Florida                       1.40120   1.58474   0.884   0.3766
poisson_data_sna$university_affiliationUniversity of Georgia                      -1.16031   2.02221  -0.574   0.5661
poisson_data_sna$university_affiliationUniversity of Hawaii Manoa                  0.80545   1.71391   0.470   0.6384
poisson_data_sna$university_affiliationUniversity of Hong Kong                     0.53517   0.89600   0.597   0.5503
poisson_data_sna$university_affiliationUniversity of Hull                         -0.73595   1.68459  -0.437   0.6622
poisson_data_sna$university_affiliationUniversity of Illinois                      3.45431   1.81023   1.908   0.0564 .
poisson_data_sna$university_affiliationUniversity of Iowa                          2.15709   2.39324   0.901   0.3674
poisson_data_sna$university_affiliationUniversity of Liechtenstein                -0.94327   1.27765  -0.738   0.4603
poisson_data_sna$university_affiliationUniversity of Ljubljana                    -2.25731   2.22828  -1.013   0.3110
poisson_data_sna$university_affiliationUniversity of Louisville                    0.44442   1.92370   0.231   0.8173
poisson_data_sna$university_affiliationUniversity of Maryland                     -0.19533   1.92238  -0.102   0.9191
poisson_data_sna$university_affiliationUniversity of Massachusetts Boston          2.13114   3.48561   0.611   0.5409
poisson_data_sna$university_affiliationUniversity of Massachusetts Dartmouth       2.39026   3.44937   0.693   0.4883
poisson_data_sna$university_affiliationUniversity of Massachusetts Lowell          0.23530   3.47402   0.068   0.9460
poisson_data_sna$university_affiliationUniversity of Michigan                      1.61183   1.57877   1.021   0.3073
poisson_data_sna$university_affiliationUniversity of Minnesota                     0.40901   1.73864   0.235   0.8140
poisson_data_sna$university_affiliationUniversity of Missouri                      0.59408   3.46172   0.172   0.8637
poisson_data_sna$university_affiliationUniversity of North Texas                   1.44448   3.31196   0.436   0.6627
poisson_data_sna$university_affiliationUniversity of Pennsylvania                  2.47748   2.78284   0.890   0.3733
poisson_data_sna$university_affiliationUniversity of Sheffield                    -0.52202   1.69870  -0.307   0.7586
poisson_data_sna$university_affiliationUniversity of South Carolina                2.58671   2.00388   1.291   0.1968
poisson_data_sna$university_affiliationUniversity of South Florida                 0.06971   1.87047   0.037   0.9703
poisson_data_sna$university_affiliationUniversity of Sussex                       -1.02561   1.76252  -0.582   0.5606
poisson_data_sna$university_affiliationUniversity of Technology                   -3.19369   3.25279  -0.982   0.3262
poisson_data_sna$university_affiliationUniversity of Texas Austin                  2.24563   2.95664   0.760   0.4475
poisson_data_sna$university_affiliationUniversity of Texas Dallas                  2.20561   1.25066   1.764   0.0778 .
poisson_data_sna$university_affiliationUniversity of Texas El Paso                 1.27932   3.11152   0.411   0.6810


poisson_data_sna$university_affiliationUniversity of Texas San Antonio                     0.85737   3.43426   0.250   0.8029
poisson_data_sna$university_affiliationUniversity of Texas Southwestern Medical Center     4.33320   2.56307   1.691   0.0909 .
poisson_data_sna$university_affiliationUniversity of Toronto                               1.66411   3.18929   0.522   0.6018
poisson_data_sna$university_affiliationUniversity of Utah                                  0.46383   1.92273   0.241   0.8094
poisson_data_sna$university_affiliationUniversity of Virginia                              2.08854   1.06412   1.963   0.0497 *
poisson_data_sna$university_affiliationUniversity of Warwick                              -0.85112   1.24612  -0.683   0.4946
poisson_data_sna$university_affiliationUniversity of Washington                            0.80172   1.20392   0.666   0.5055
poisson_data_sna$university_affiliationUniversity of Wisconsin Milwaukee                   1.82824   3.69499   0.495   0.6207
poisson_data_sna$university_affiliationUniversity of Wollongong                           -1.24940   1.65464  -0.755   0.4502
poisson_data_sna$university_affiliationUniversity of Zurich                               -0.72395   3.30734  -0.219   0.8267
poisson_data_sna$university_affiliationVirginia Polytech                                   2.06802   1.42371   1.453   0.1463
poisson_data_sna$university_affiliationWashington State University                         1.23006   1.63248   0.753   0.4512
poisson_data_sna$university_affiliationYonsei University                                  -0.98199   1.04815  -0.937   0.3488
poisson_data_sna$university_affiliationYork University                                          NA        NA      NA       NA
poisson_data_sna$dept_affiliationDecision Sciences                                              NA        NA      NA       NA
poisson_data_sna$dept_affiliationEconomics                                                 3.78786   2.19489   1.726   0.0844 .
poisson_data_sna$dept_affiliationEngineering Management                                         NA        NA      NA       NA
poisson_data_sna$dept_affiliationFinance                                                   1.63625   1.40999   1.160   0.2459
poisson_data_sna$dept_affiliationIE                                                             NA        NA      NA       NA
poisson_data_sna$dept_affiliationIndustry                                                 -0.99627   1.74913  -0.570   0.5690
poisson_data_sna$dept_affiliationInnovation and Entrepreneurship                           2.30768   2.38516   0.968   0.3333
poisson_data_sna$dept_affiliationInternational Business                                         NA        NA      NA       NA
poisson_data_sna$dept_affiliationIS                                                        1.61978   0.80234   2.019   0.0435 *
poisson_data_sna$dept_affiliationIT                                                        3.15049   1.58478   1.988   0.0468 *
poisson_data_sna$dept_affiliationIT Management                                             2.76386   1.82545   1.514   0.1300
poisson_data_sna$dept_affiliationMarketing                                                 1.73826   1.65356   1.051   0.2932
poisson_data_sna$dept_affiliationOM                                                        1.57238   1.56400   1.005   0.3147
poisson_data_sna$dept_affiliationPsychology                                                     NA        NA      NA       NA
poisson_data_sna$dept_affiliationStatistics                                                     NA        NA      NA       NA
poisson_data_sna$dept_affiliationTechnology Management                                     1.99678   2.74937   0.726   0.4677
poisson_data_sna$experience                                                               -0.01316   0.01915  -0.687   0.4918
poisson_data_sna$current_carnegie16                                                        0.64389   2.75390   0.234   0.8151
poisson_data_sna$current_carnegie17                                                             NA        NA      NA       NA
poisson_data_sna$current_carnegie18                                                             NA        NA      NA       NA
poisson_data_sna$current_carnegie21                                                             NA        NA      NA       NA
poisson_data_sna$current_carnegie25                                                             NA        NA      NA       NA
poisson_data_sna$current_carnegie99                                                       -0.74720   1.22749  -0.609   0.5427
```

```
poisson_data_sna$phd_granting_instAthens University of Economics and Business          -0.29455    1.66436  -0.177   0.8595
poisson_data_sna$phd_granting_instAustralian Graduate School of Management              1.24940    1.65464   0.755   0.4502
poisson_data_sna$phd_granting_instCambridge University                                       NA         NA      NA       NA
poisson_data_sna$phd_granting_instCardiff University                                   -1.29064    2.82763  -0.456   0.6481
poisson_data_sna$phd_granting_instCase Western Reserve University                       1.19374    1.14988   1.038   0.2992
poisson_data_sna$phd_granting_instChinese University of HongKong                             NA         NA      NA       NA
poisson_data_sna$phd_granting_instClemson University                                    0.49568    1.25798   0.394   0.6936
poisson_data_sna$phd_granting_instCMU                                                  -0.55317    2.63922  -0.210   0.8340
poisson_data_sna$phd_granting_instColumbia University                                   0.15797    1.43276   0.110   0.9122
poisson_data_sna$phd_granting_instCopenhagen Business School                                 NA         NA      NA       NA
poisson_data_sna$phd_granting_instCranfield University                                       NA         NA      NA       NA
poisson_data_sna$phd_granting_instCurtin University                                          NA         NA      NA       NA
poisson_data_sna$phd_granting_instEcole des Hautes Etudes en Sciences Sociales         -0.66798    3.08656  -0.216   0.8287
poisson_data_sna$phd_granting_instEcole Superiere des Affaires                          0.42461    1.85176   0.229   0.8186
poisson_data_sna$phd_granting_instErasmus University                                         NA         NA      NA       NA
poisson_data_sna$phd_granting_instFlorida Inernational University                            NA         NA      NA       NA
poisson_data_sna$phd_granting_instFlorida International University                           NA         NA      NA       NA
poisson_data_sna$phd_granting_instFlorida State University                                   NA         NA      NA       NA
poisson_data_sna$phd_granting_instFundação Getúlio Vargas                                    NA         NA      NA       NA
poisson_data_sna$phd_granting_instGeorgia State University                                   NA         NA      NA       NA
poisson_data_sna$phd_granting_instHKUST                                                 1.62309    1.40733   1.153   0.2488
poisson_data_sna$phd_granting_instIIT India                                                  NA         NA      NA       NA
poisson_data_sna$phd_granting_instIndiana University                                   -1.07067    2.23084  -0.480   0.6313
poisson_data_sna$phd_granting_instJohns Hopkins University                              1.46611    1.72319   0.851   0.3949
poisson_data_sna$phd_granting_instKatholieke University Leuven                               NA         NA      NA       NA
poisson_data_sna$phd_granting_instKent State University                                 0.57554    1.68638   0.341   0.7329
poisson_data_sna$phd_granting_instLSE                                                        NA         NA      NA       NA
poisson_data_sna$phd_granting_instMcGill University                                          NA         NA      NA       NA
poisson_data_sna$phd_granting_instMedical College of Wisconsin                          0.41369    2.18550   0.189   0.8499
poisson_data_sna$phd_granting_instMIT                                                  -1.02362    1.58420  -0.646   0.5182
poisson_data_sna$phd_granting_instMonash University                                          NA         NA      NA       NA
poisson_data_sna$phd_granting_instNorthwestern University                                    NA         NA      NA       NA
poisson_data_sna$phd_granting_instNova Information Management School                         NA         NA      NA       NA
poisson_data_sna$phd_granting_instNUS                                                  -1.26257    1.65007  -0.765   0.4442
poisson_data_sna$phd_granting_instNYU                                                   1.34465    0.91723   1.466   0.1427
poisson_data_sna$phd_granting_instPennsylvinia State University                        -0.91386    2.47093  -0.370   0.7115
```

```
poisson_data_sna$phd_granting_instPurdue University                                    -1.22902    1.58017  -0.778   0.4367
poisson_data_sna$phd_granting_instQueensland University of Technology                   1.06243    0.81789   1.299   0.1939
poisson_data_sna$phd_granting_instRMIT                                                  0.18955    2.01383   0.094   0.9250
poisson_data_sna$phd_granting_instStevens Institute of Technology                            NA         NA      NA       NA
poisson_data_sna$phd_granting_instSyracuse University                                   1.23774    1.96189   0.631   0.5281
poisson_data_sna$phd_granting_instTechnion Israel Institute of Technology                    NA         NA      NA       NA
poisson_data_sna$phd_granting_instTexas Tech                                            0.97078    1.36951   0.709   0.4784
poisson_data_sna$phd_granting_instUBC                                                  -2.06007    1.64876  -1.249   0.2115
poisson_data_sna$phd_granting_instUC Berkeley                                           0.10863    1.16613   0.093   0.9258
poisson_data_sna$phd_granting_instUC Irvine                                                  NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversidade do Vale do Rio dos Sinos                      NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversita Cattolica del Sacro Cuore                  1.86530    2.17129   0.859   0.3903
poisson_data_sna$phd_granting_instUniversity of Arizona                                 1.41436    0.71043   1.991   0.0465 *
poisson_data_sna$phd_granting_instUniversity of Bradford                                     NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Cambridge                               1.17760    1.82935   0.644   0.5198
poisson_data_sna$phd_granting_instUniversity of Cape Town                                    NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Cologne                                      NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Georgia                                -1.27605    1.28545  -0.993   0.3209
poisson_data_sna$phd_granting_instUniversity of Houston                                      NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Iowa                                    0.40772    2.25319   0.181   0.8564
poisson_data_sna$phd_granting_instUniversity of Kentucky                                     NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Leeds                                        NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Liverpool                                    NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Ljublijana                                   NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Maryland                                     NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Michigan                                     NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Minnesota                               2.73265    1.21101   2.256   0.0240 *
poisson_data_sna$phd_granting_instUniversity of Missouri                               -0.53783    2.50108  -0.215   0.8297
poisson_data_sna$phd_granting_instUniversity of Muenster                                1.16443    1.15866   1.005   0.3149
poisson_data_sna$phd_granting_instUniversity of Munich                                       NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Oklahama                               -1.10871    3.60459  -0.308   0.7584
poisson_data_sna$phd_granting_instUniversity of Oulu                                    1.18246    1.56625   0.755   0.4503
poisson_data_sna$phd_granting_instUniversity of Pittsburgh                              2.33001    1.13338   2.056   0.0398 *
poisson_data_sna$phd_granting_instUniversity of Rochester                               0.60492    0.73154   0.827   0.4083
poisson_data_sna$phd_granting_instUniversity of Sourthern California                         NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Strathclyde                                  NA         NA      NA       NA
```

```
poisson_data_sna$phd_granting_instUniversity of Toronto                        NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Utah                      0.26252    1.26476   0.208   0.8356
poisson_data_sna$phd_granting_instUniversity of Washington                0.04056    1.19055   0.034   0.9728
poisson_data_sna$phd_granting_instUniversity of Wollongong                     NA         NA      NA       NA
poisson_data_sna$phd_granting_instUppsala University                      0.40810    1.53376   0.266   0.7902
poisson_data_sna$phd_granting_instUT Arlington                                 NA         NA      NA       NA
poisson_data_sna$phd_granting_instUT Austin                               0.54093    0.51687   1.047   0.2953
poisson_data_sna$phd_granting_instUT Dallas                               1.50512    1.56942   0.959   0.3375
poisson_data_sna$phd_granting_instVrije University                             NA         NA      NA       NA
poisson_data_sna$phd_granting_instWashington State University                  NA         NA      NA       NA
poisson_data_sna$phd_granting_instYonsei University                            NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 143.0205  on 183  degrees of freedom
Residual deviance:   7.8426  on  15  degrees of freedom
AIC: 762.99

Number of Fisher Scoring iterations: 4
```

Figure A.8. Poisson regression with node degree as dependent variable (with Carnegie classification variable)

```
Call:
glm(formula = poisson_data_sna$degree ~ poisson_data_sna$gender +
    poisson_data_sna$continental_affiliation + poisson_data_sna$university_affiliation +
    poisson_data_sna$dept_affiliation + poisson_data_sna$experience +
    poisson_data_sna$phd_granting_inst, family = poisson(link = "log"),
    data = poisson_data_sna)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9567   0.0000   0.0000   0.0000   0.7410

Coefficients: (48 not defined because of singularities)
                                                                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                                                                        -2.368829   1.549226  -1.529  0.12625
poisson_data_sna$genderMale                                                        -0.264495   1.040908  -0.254  0.79942
poisson_data_sna$continental_affiliationNon_american                                1.905480   1.345517   1.416  0.15673
poisson_data_sna$university_affiliationAT&T Lab                                     2.591561   1.891123   1.370  0.17057
poisson_data_sna$university_affiliationAuburn University                            0.921634   1.721733   0.535  0.59245
poisson_data_sna$university_affiliationAustralian National University               0.562128   1.669231   0.337  0.73630
poisson_data_sna$university_affiliationBentley University                          -0.343107   1.731270  -0.198  0.84290
poisson_data_sna$university_affiliationBlue Slate Solutions                         0.843023   1.707789   0.494  0.62156
poisson_data_sna$university_affiliationBrown University                             1.572198   1.779480   0.884  0.37696
poisson_data_sna$university_affiliationBrunel University                           -1.850566   1.593497  -1.161  0.24551
poisson_data_sna$university_affiliationBYU                                          0.124006   1.319029   0.094  0.92510
poisson_data_sna$university_affiliationCalifornia State University at Stanislaus    2.314118   2.191745   1.056  0.29104
poisson_data_sna$university_affiliationCardiff University                          -1.167273   2.100422  -0.556  0.57839
poisson_data_sna$university_affiliationChinese University of Hong Kong              1.094697   1.481025   0.739  0.45982
poisson_data_sna$university_affiliationCity University of Hong Kong                -0.735733   0.915308  -0.804  0.42151
poisson_data_sna$university_affiliationCMU                                          0.778875   1.579844   0.493  0.62201
poisson_data_sna$university_affiliationCollege of Business Administration Kuwait   -0.983846   1.597452  -0.616  0.53797
poisson_data_sna$university_affiliationColorado State University                    1.251639   1.516199   0.826  0.40908
poisson_data_sna$university_affiliationCopenhagen Business School                  -1.028483   1.419244  -0.725  0.46865
poisson_data_sna$university_affiliationCoriolis Labs                                4.223698   2.466955   1.712  0.08688 .
poisson_data_sna$university_affiliationCurtin University of Technology              0.379261   0.913667   0.415  0.67807
poisson_data_sna$university_affiliationDeloitte MCS Ltd                             4.223698   2.466955   1.712  0.08688 .
poisson_data_sna$university_affiliationDrexel University                           -0.104815   1.422411  -0.074  0.94126
poisson_data_sna$university_affiliationEmory University                             1.341003   1.423818   0.942  0.34628
poisson_data_sna$university_affiliationErasmus University                          -0.758657   1.228447  -0.618  0.53686
poisson_data_sna$university_affiliationESAN University                              0.418567   0.913070   0.458  0.64665

poisson_data_sna$university_affiliationESC Rennes School of Business                0.094171   2.088056   0.045  0.96403
poisson_data_sna$university_affiliationESSEC Business School                        0.663334   2.346672   0.283  0.77743
poisson_data_sna$university_affiliationFlorida State University                     1.943460   1.626600   1.195  0.23217
poisson_data_sna$university_affiliationFundação Getúlio Vargas                     -1.075560   1.569717  -0.685  0.49322
poisson_data_sna$university_affiliationGeorgia State University                     0.933020   1.618596   0.576  0.56432
poisson_data_sna$university_affiliationHealthPartners Institute for Education and Research  2.921566   1.696293   1.722  0.08501 .
poisson_data_sna$university_affiliationHEC Montreal                                 0.633673   1.769406   0.358  0.72025
poisson_data_sna$university_affiliationIMT Ghaziabad                               -1.929177   1.602773  -1.204  0.22873
poisson_data_sna$university_affiliationIndiana University                           2.381136   2.581644   0.922  0.35635
poisson_data_sna$university_affiliationINSEAD                                      -0.837268   1.882541  -0.445  0.65650
poisson_data_sna$university_affiliationJohns Hopkins University                    -1.728813   1.781803  -0.970  0.33192
poisson_data_sna$university_affiliationKeele University                            -1.131247   1.938985  -0.583  0.55961
poisson_data_sna$university_affiliationKonyang University                           0.379261   0.913667   0.415  0.67807
poisson_data_sna$university_affiliationKorea Advanced Institute of Science and Technology -0.627638   1.228447  -0.511  0.60941
poisson_data_sna$university_affiliationLondon School of Economics                  -1.922778   1.694008  -1.135  0.25636
poisson_data_sna$university_affiliationLouisiana State University                   2.167518   1.362679   1.591  0.11169
poisson_data_sna$university_affiliationLudwig-Maximilian-University Munich         -2.012753   2.368937  -0.850  0.39552
poisson_data_sna$university_affiliationMiami University                             1.186129   1.523730   0.778  0.43631
poisson_data_sna$university_affiliationMichigan State University                    1.317148   1.514638   0.870  0.38451
poisson_data_sna$university_affiliationMillsaps College                             1.068212   1.552012   0.688  0.49128
poisson_data_sna$university_affiliationMonash University                            0.379261   0.913667   0.415  0.67807
poisson_data_sna$university_affiliationNUS                                         -0.026204   1.000726  -0.026  0.97911
poisson_data_sna$university_affiliationNYU                                          1.666864   1.243424   1.341  0.18007
poisson_data_sna$university_affiliationOklahama State                               2.328728   2.602018   0.895  0.37080
poisson_data_sna$university_affiliationPennsylvania State University                1.251639   2.073369   0.604  0.54606
poisson_data_sna$university_affiliationPurdue University                            2.129344   2.356816   0.903  0.36627
poisson_data_sna$university_affiliationRMIT University                             -0.719351   1.225338  -0.587  0.55716
poisson_data_sna$university_affiliationSan Diego State University                   0.026204   1.414727   0.019  0.98522
poisson_data_sna$university_affiliationShells Chemicals Seraya Pvt Ltd              3.157943   2.724324   1.159  0.24639
poisson_data_sna$university_affiliationSimon Fraser University                     -0.758657   1.228447  -0.618  0.53686
poisson_data_sna$university_affiliationStevens Institute of Technology              3.184502   1.933951   1.647  0.09963 .
poisson_data_sna$university_affiliationTel-Aviv University                          0.672978   1.786003   0.377  0.70632
poisson_data_sna$university_affiliationTennessee Technological University           2.908464   1.695329   1.716  0.08624 .
poisson_data_sna$university_affiliationTexas Christian University                   2.194854   1.854169   1.184  0.23652
poisson_data_sna$university_affiliationTexas Tech                                   0.765773   1.579802   0.485  0.62787
poisson_data_sna$university_affiliationThe Hong Kong Polytechnic University        -0.797962   1.234201  -0.647  0.51793
```

```
poisson_data_sna$university_affiliationThe University of Hong Kong                  -1.264741  1.138622  -1.111  0.26667
poisson_data_sna$university_affiliationThree Gorges University                     -1.804376  2.119334  -0.851  0.39455
poisson_data_sna$university_affiliationTsinghua University                         -2.121279  2.768131  -0.766  0.44348
poisson_data_sna$university_affiliationUniversidade do Vale do Rio dos Sinos       -0.313886  0.913666  -0.344  0.73119
poisson_data_sna$university_affiliationUniversidade NOVA ISEGI                      0.950576  1.447455   0.657  0.51136
poisson_data_sna$university_affiliationUniversity of Amsterdam                     -0.922278  1.403519  -0.657  0.51110
poisson_data_sna$university_affiliationUniversity of Antwerp                        0.950576  1.447455   0.657  0.51136
poisson_data_sna$university_affiliationUniversity of Arizona                        1.113629  1.127622   0.988  0.32335
poisson_data_sna$university_affiliationUniversity of Calgary                        0.222039  0.951075   0.233  0.81540
poisson_data_sna$university_affiliationUniversity of California Irvine              1.199231  1.521749   0.788  0.43066
poisson_data_sna$university_affiliationUniversity of Cincinnati                     1.353112  1.528429   0.885  0.37600
poisson_data_sna$university_affiliationUniversity of Cologne                        0.031980  2.055180   0.016  0.98758
poisson_data_sna$university_affiliationUniversity of Florida                        1.173028  1.525946   0.769  0.44206
poisson_data_sna$university_affiliationUniversity of Georgia                       -1.755017  1.782681  -0.984  0.32488
poisson_data_sna$university_affiliationUniversity of Hawaii Manoa                   0.829921  1.706200   0.486  0.62667
poisson_data_sna$university_affiliationUniversity of Hong Kong                      0.535924  0.895722   0.598  0.54963
poisson_data_sna$university_affiliationUniversity of Hull                          -0.735733  1.684574  -0.437  0.66229
poisson_data_sna$university_affiliationUniversity of Illinois                       3.226265  1.759037   1.834  0.06664 .
poisson_data_sna$university_affiliationUniversity of Iowa                           2.581917  1.383873   1.866  0.06208 .
poisson_data_sna$university_affiliationUniversity of Liechtenstein                 -0.942083  1.277167  -0.738  0.46074
poisson_data_sna$university_affiliationUniversity of Ljubljana                     -1.775243  2.061413  -0.861  0.38914
poisson_data_sna$university_affiliationUniversity of Louisville                    -0.155308  1.660324  -0.094  0.92547
poisson_data_sna$university_affiliationUniversity of Maryland                       0.054626  1.522031   0.036  0.97137
poisson_data_sna$university_affiliationUniversity of Massachusetts Boston           2.287914  2.199367   1.040  0.29822
poisson_data_sna$university_affiliationUniversity of Massachusetts Dartmouth        2.856056  1.693616   1.686  0.09173 .
poisson_data_sna$university_affiliationUniversity of Massachusetts Lowell           0.604333  1.944022   0.311  0.75590
poisson_data_sna$university_affiliationUniversity of Michigan                       1.382658  1.519067   0.910  0.36272
poisson_data_sna$university_affiliationUniversity of Minnesota                     -0.175104  1.453720  -0.120  0.90412
poisson_data_sna$university_affiliationUniversity of Missouri                       0.711606  2.033768   0.350  0.72642
poisson_data_sna$university_affiliationUniversity of North Texas                    2.111960  1.544845   1.367  0.17159
poisson_data_sna$university_affiliationUniversity of Pennsylvania                   1.996869  2.646964   0.754  0.45061
poisson_data_sna$university_affiliationUniversity of Sheffield                     -0.568694  1.703112  -0.334  0.73844
poisson_data_sna$university_affiliationUniversity of South Carolina                 2.059154  1.800222   1.144  0.25269
poisson_data_sna$university_affiliationUniversity of South Florida                 -0.526380  1.601966  -0.329  0.74247
poisson_data_sna$university_affiliationUniversity of Sussex                        -0.774216  1.698921  -0.456  0.64860
poisson_data_sna$university_affiliationUniversity of Technology                    -2.460675  2.980829  -0.826  0.40909
poisson_data_sna$university_affiliationUniversity of Texas Austin                   2.458607  2.159193   1.139  0.25484
poisson_data_sna$university_affiliationUniversity of Texas Dallas                   1.971011  1.173606   1.679  0.09306 .

poisson_data_sna$university_affiliationUniversity of Texas El Paso                  1.695850  1.408013   1.204  0.22842
poisson_data_sna$university_affiliationUniversity of Texas San Antonio              0.973644  1.975524   0.493  0.62212
poisson_data_sna$university_affiliationUniversity of Texas Southwestern Medical Center  4.525042  2.557649   1.769  0.07686 .
poisson_data_sna$university_affiliationUniversity of Toronto                        2.155829  2.315411   0.931  0.35181
poisson_data_sna$university_affiliationUniversity of Utah                           0.487183  1.910325   0.255  0.79870
poisson_data_sna$university_affiliationUniversity of Virginia                       2.037938  1.062785   1.918  0.05517 .
poisson_data_sna$university_affiliationUniversity of Warwick                       -0.850370  1.245920  -0.683  0.49491
poisson_data_sna$university_affiliationUniversity of Washington                     0.766998  1.195459   0.642  0.52114
poisson_data_sna$university_affiliationUniversity of Wisconsin Milwaukee            3.289690  2.983951   1.102  0.27026
poisson_data_sna$university_affiliationUniversity of Wollongong                    -0.996948  1.592835  -0.626  0.53138
poisson_data_sna$university_affiliationUniversity of Zurich                         0.718025  1.934759   0.371  0.71055
poisson_data_sna$university_affiliationVirginia Polytech                            1.839971  1.358017   1.355  0.17545
poisson_data_sna$university_affiliationWashington State University                  1.002703  1.575596   0.636  0.52452
poisson_data_sna$university_affiliationYonsei University                           -0.996517  1.048527  -0.950  0.34191
poisson_data_sna$university_affiliationYork University                                    NA        NA      NA       NA
poisson_data_sna$dept_affiliationDecision Sciences                                        NA        NA      NA       NA
poisson_data_sna$dept_affiliationEconomics                                          3.356386  2.036461   1.648  0.09932 .
poisson_data_sna$dept_affiliationEngineering Management                                   NA        NA      NA       NA
poisson_data_sna$dept_affiliationFinance                                            1.686308  1.396502   1.208  0.22723
poisson_data_sna$dept_affiliationIE                                                       NA        NA      NA       NA
poisson_data_sna$dept_affiliationIndustry                                          -1.367642  1.672176  -0.818  0.41342
poisson_data_sna$dept_affiliationInnovation and Entrepreneurship                    2.104747  2.356689   0.893  0.37181
poisson_data_sna$dept_affiliationInternational Business                                   NA        NA      NA       NA
poisson_data_sna$dept_affiliationIS                                                 1.669927  0.778756   2.144  0.03200 *
poisson_data_sna$dept_affiliationIT                                                 2.801142  1.422016   1.970  0.04886 *
poisson_data_sna$dept_affiliationIT Management                                      2.562060  1.784485   1.436  0.15108
poisson_data_sna$dept_affiliationMarketing                                          1.787844  1.641169   1.089  0.27599
poisson_data_sna$dept_affiliationOM                                                 1.921718  1.479594   1.299  0.19401
poisson_data_sna$dept_affiliationPsychology                                               NA        NA      NA       NA
poisson_data_sna$dept_affiliationStatistics                                               NA        NA      NA       NA
poisson_data_sna$dept_affiliationTechnology Management                              0.706508  1.942871   0.364  0.71613
poisson_data_sna$experience                                                        -0.013102  0.019061  -0.687  0.49185
poisson_data_sna$phd_granting_instAthens University of Economics and Business      -0.007771  1.612063  -0.005  0.99615
poisson_data_sna$phd_granting_instAustralian Graduate School of Management          0.996948  1.592835   0.626  0.53138
poisson_data_sna$phd_granting_instCambridge University                                    NA        NA      NA       NA
poisson_data_sna$phd_granting_instCardiff University                               -0.809885  2.694471  -0.301  0.76374
poisson_data_sna$phd_granting_instCase Western Reserve University                   1.564170  1.024596   1.527  0.12686
poisson_data_sna$phd_granting_instChinese University of HongKong                          NA        NA      NA       NA
```

```
poisson_data_sna$phd_granting_instClemson University                       0.496619   1.257675   0.395  0.69294
poisson_data_sna$phd_granting_instCMU                                     -0.995623   1.762615  -0.565  0.57217
poisson_data_sna$phd_granting_instColumbia University                      0.157223   1.432591   0.110  0.91261
poisson_data_sna$phd_granting_instCopenhagen Business School                     NA         NA      NA       NA
poisson_data_sna$phd_granting_instCranfield University                           NA         NA      NA       NA
poisson_data_sna$phd_granting_instCurtin University                              NA         NA      NA       NA
poisson_data_sna$phd_granting_instEcole des Hautes Etudes en Sciences Sociales  0.924438   1.855980   0.498  0.61842
poisson_data_sna$phd_granting_instEcole Superiere des Affaires             0.172782   1.793630   0.096  0.92326
poisson_data_sna$phd_granting_instErasmus University                             NA         NA      NA       NA
poisson_data_sna$phd_granting_instFlorida Inernational University                NA         NA      NA       NA
poisson_data_sna$phd_granting_instFlorida International University               NA         NA      NA       NA
poisson_data_sna$phd_granting_instFlorida State University                       NA         NA      NA       NA
poisson_data_sna$phd_granting_instFundação Getúlio Vargas                        NA         NA      NA       NA
poisson_data_sna$phd_granting_instGeorgia State University                       NA         NA      NA       NA
poisson_data_sna$phd_granting_instHKUST                                    1.673207   1.393949   1.200  0.23001
poisson_data_sna$phd_granting_instIIT India                                      NA         NA      NA       NA
poisson_data_sna$phd_granting_instIndiana University                      -1.168802   2.074541  -0.563  0.57316
poisson_data_sna$phd_granting_instJohns Hopkins University                 1.819800   1.643468   1.107  0.26817
poisson_data_sna$phd_granting_instKatholieke University Leuven                   NA         NA      NA       NA
poisson_data_sna$phd_granting_instKent State University                    0.371027   1.636317   0.227  0.82062
poisson_data_sna$phd_granting_instLSE                                            NA         NA      NA       NA
poisson_data_sna$phd_granting_instMcGill University                              NA         NA      NA       NA
poisson_data_sna$phd_granting_instMedical College of Wisconsin             1.037637   1.967600   0.527  0.59794
poisson_data_sna$phd_granting_instMIT                                     -0.771434   1.517959  -0.508  0.61131
poisson_data_sna$phd_granting_instMonash University                              NA         NA      NA       NA
poisson_data_sna$phd_granting_instNorthwestern University                        NA         NA      NA       NA
poisson_data_sna$phd_granting_instNova Information Management School             NA         NA      NA       NA
poisson_data_sna$phd_granting_instNUS                                     -1.010050   1.588434  -0.636  0.52486
poisson_data_sna$phd_granting_instNYU                                      1.712510   0.765454   2.237  0.02527 *
poisson_data_sna$phd_granting_instPennsylvinia State University           -1.566112   1.446360  -1.083  0.27890
poisson_data_sna$phd_granting_instPurdue University                       -0.970765   1.511758  -0.642  0.52078
poisson_data_sna$phd_granting_instQueensland University of Technology      1.062586   0.817884   1.299  0.19388
poisson_data_sna$phd_granting_instRMIT                                     0.441754   1.962243   0.225  0.82188
poisson_data_sna$phd_granting_instStevens Institute of Technology                NA         NA      NA       NA
poisson_data_sna$phd_granting_instSyracuse University                      0.926146   1.405264   0.659  0.50986
poisson_data_sna$phd_granting_instTechnion Israel Institute of Technology        NA         NA      NA       NA
poisson_data_sna$phd_granting_instTexas Tech                               0.995586   1.368575   0.727  0.46694
poisson_data_sna$phd_granting_instUBC                                     -1.838902   1.407209  -1.307  0.19129
poisson_data_sna$phd_granting_instUC Berkeley                              0.108095   1.166016   0.093  0.92614

poisson_data_sna$phd_granting_instUC Irvine                                      NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversidade do Vale do Rio dos Sinos          NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversita Cattolica del Sacro Cuore     1.385115   1.992464   0.695  0.48694
poisson_data_sna$phd_granting_instUniversity of Arizona                    1.186129   0.567233   2.091  0.03652 *
poisson_data_sna$phd_granting_instUniversity of Bradford                         NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Cambridge                  1.177224   1.829320   0.644  0.51988
poisson_data_sna$phd_granting_instUniversity of Cape Town                        NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Cologne                          NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Georgia                   -1.074234   1.012343  -1.061  0.28863
poisson_data_sna$phd_granting_instUniversity of Houston                          NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Iowa                      -0.237396   1.107273  -0.214  0.83024
poisson_data_sna$phd_granting_instUniversity of Kentucky                         NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Leeds                            NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Liverpool                        NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Ljublijana                       NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Maryland                         NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Michigan                         NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Minnesota                  3.098369   1.101071   2.814  0.00489 **
poisson_data_sna$phd_granting_instUniversity of Missouri                   0.013500   2.340701   0.006  0.99540
poisson_data_sna$phd_granting_instUniversity of Muenster                   1.164122   1.158627   1.005  0.31502
poisson_data_sna$phd_granting_instUniversity of Munich                           NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Oklahama                  -0.803270   2.948184  -0.272  0.78527
poisson_data_sna$phd_granting_instUniversity of Oulu                       0.918003   1.487000   0.617  0.53700
poisson_data_sna$phd_granting_instUniversity of Pittsburgh                 2.379518   1.114933   2.134  0.03282 *
poisson_data_sna$phd_granting_instUniversity of Rochester                  0.604776   0.731518   0.827  0.40838
poisson_data_sna$phd_granting_instUniversity of Southern California              NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Strathclyde                      NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Toronto                          NA         NA      NA       NA
poisson_data_sna$phd_granting_instUniversity of Utah                       0.277166   1.265117   0.219  0.82659
poisson_data_sna$phd_granting_instUniversity of Washington                 0.125722   1.122522   0.112  0.91082
poisson_data_sna$phd_granting_instUniversity of Wollongong                       NA         NA      NA       NA
poisson_data_sna$phd_granting_instUppsala University                       0.406159   1.532692   0.265  0.79101
poisson_data_sna$phd_granting_instUT Arlington                                   NA         NA      NA       NA
poisson_data_sna$phd_granting_instUT Austin                                0.564477   0.513992   1.098  0.27211
poisson_data_sna$phd_granting_instUT Dallas                                1.877362   1.485585   1.264  0.20633
poisson_data_sna$phd_granting_instVrije University                               NA         NA      NA       NA
poisson_data_sna$phd_granting_instWashington State University                    NA         NA      NA       NA
poisson_data_sna$phd_granting_instYonsei University                              NA         NA      NA       NA
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 143.0205  on 183  degrees of freedom
Residual deviance:   8.2536  on  17  degrees of freedom
AIC: 759.4

Number of Fisher Scoring iterations: 4
```

Figure A.9. Poisson regression with node degree as dependent variable (without Carnegie classification variable)

Chapter 3 Breaching Together: A Data Science Approach on Firm's Correlated Risk in Information Security

3.1 Abstract

This study develops a data science approach to measuring relatedness of business firms, and it aims to assess the extent to which such a measure of business relatedness correlates to firms' risk of information security breaches. I analyze unstructured textual descriptions from Securities and Exchange Commission (SEC) 10-K filing reports of 33 public firms that were breached together (on the same date and by the same breach) in the last 10 years, 2008–2017. Specifically, I employ the text analysis technique of topic modeling to measure breach proximity based on business descriptions, security risk factors, and internal control reporting sections of the 10-K filings. In addition, the Quadratic Assignment Procedure (QAP), a well-known technique in social network analysis, tests for significance in statistical relationships among the various similarity matrices. Results show that dyadic relationships between public firms based on textual descriptions from their 10-K filings are significantly correlated with the dyads based on information security breaches for these firms. Along with SEC 10-K filings for the 33 firms, I also consider other firm-level characteristics such as geographical proximity, industry type, size, revenue and research and development (R&D) expenditure. Security risk factors, industry classification based on 2-digit SIC code, and R&D intensity are significantly correlated with co-occurrence of breaches among firms.

3.2 Introduction

An information security breach is unauthorized access or acquisition of data (digital or analog) that compromises “the security, confidentiality, or integrity” of proprietary or personal information maintained by an individual or an organization (Faulkner, 2007). Examples of information security breaches include the theft of a disk or portable device with classified data; consumer data obtained by hackers; or the theft of proprietary information by insiders. As per one public source, 8064 data breaches have occurred since 2005, and those have compromised more than 10 billion user records.[16] Although businesses have increased annual security spending,[17] many still suffer from heavy financial loss due to security breaches. For example, IBM recently released a study showing that the average total cost of data breaches across the globe is $3.62 million for the year 2017, and the average cost per lost record in USD is $141.[18] Another recent security breach at a credit reporting agency exposed the sensitive personal information of about 143 million US consumers.[19]

This chapter empirically investigates correlated risks and failures in information security at the firm level. Cybersecurity risk is risk that arises from malicious electronic or non-electronic events affecting the information technology resources of firms, and they often result in disruption of business and financial loss (Biener, Eling, & Wirfs, 2015; Mukhopadhyay, Chatterjee, Saha, Mahanti, & Sadhukhan, 2013). From a technological standpoint, firms often share correlated risks and vulnerabilities to breach due to the usage of common security technologies and the connectivity of computer networks (Chen, Kataria, & Krishnan, 2011; Öğüt, Raghunathan, & Menon, 2011). The role of correlated risks has been widely investigated by the cybersecurity insurance community (Baer & Parkinson, 2007; Böhme & Kataria, 2006; Böhme & Schwartz, 2010; Mukhopadhyay et al., 2013). Historically, correlated risks and failures are investigated

[16] https://www.privacyrights.org/data-breaches
[17] https://www.gartner.com/newsroom/id/3836563, https://www.sans.org/reading-room/whitepapers/analyst/security-spending-trends-36697
[18] https://www.ibm.com/security/data-breach
[19] https://www.consumer.ftc.gov/blog/2017/09/equifax-data-breach-what-do

either within a firm among multiple systems on its own internal networks or across firms on their respective external networks (Chen et al., 2011; Mukhopadhyay et al., 2013).

This study examines the likelihood of organizations to make public announcements of data breaches on the same day. I empirically investigate the relatedness of businesses and their likelihoods of being breached together. My research question concerns the reasoning that business proximity between firms, as determined by the relatedness of certain organizational attributes, is a likely predictor of their being concurrently breached. I aim to define the characteristics or attributes of firms, the similarities of which may cause them to be more vulnerable to risk of breaching together.

Business proximity is the relatedness of businesses in terms of products, operative markets, and/or technological resources (Shi et al., 2015). As noted in the literature review section below, research on shared vulnerabilities derived from interconnected computer networks and homogenous software stacks is extant. However, little or no empirical investigation exists on the correlation between business relatedness and cybersecurity failures. This research helps fill this void.

This chapter differs from previous IS studies on cybersecurity breaches in several ways. First, prior studies were mostly confined to studying breaches either at the individual level or at the organizational level. For instance, some studies examined the organizational effects of security breaches on firms' financial performances (Acquisti, Friedman, & Telang, 2006; Avery & Ranganathan, 2016; Campbell, Gordon, Loeb, & Zhou, 2003; Ko & Dorantes, 2006). Others explore the effects of compliance policies on security breaches within organizations (Ernest Chang & Lin, 2007; Kraemer, Carayon, & Clem, 2009). Still other studied investigate the economics of information security investments (Gordon & Loeb, 2002; Huang, Hu, & Behara, 2006). Two recent studies analyzed the relationship between security investments and breaches for the healthcare industry (Angst, Block, D'arcy, & Kelley, 2017; Kwon & Johnson, 2014). None of the previous studies attempt to understand security breaches at the dyadic level—that is, how the ties between firms based on similarity impact their being breached together. My objective in this chapter is to perform such dyadic analyses in order to identify the business attributes that lead to the greatest increases in potential for a cybersecurity breach attempt. These conclusions will help business leaders in firms with those attributes exercise caution and think ahead of attackers.

This study is also unique from a social network perspective. I consider the network of firms that have been breached together as a negative link network or negative social network. Previous studies on social network analysis emphasize the positive aspects of social networks, such as knowledge transfer and social capital increases (Coleman, 2000; Inkpen & Tsang, 2005; Maurer, Bartsch, & Ebers, 2011; Nahapiet & Ghoshal, 1998; Tsai & Ghoshal, 1998); the advantages of strong and weak networks (Constant, Sproull, & Kiesler, 1996; M. Granovetter, 1983; M. S. Granovetter, 1977; Ruef, 2002); and the small world phenomenon (Latora & Marchiori, 2001). Although four social network studies examine negative relationships at the individual level (Labianca & Brass, 2006; Leskovec, Huttenlocher, & Kleinberg, 2010; Rook, 1984) or the negative aspects of social capital (Portes, 2014), the literature is scant outside of these four articles. Hence, our analysis of negative dyadic relationships with non-intentional links (because no business strives to be part of a network based on information security breaches) between firms in the context of security breaches is unique.

3.3 Background Literature and Variables Used

In this exploratory study, we investigate whether firms with similar organizational characteristics are more likely to be breached together. The firm characteristics (specifically, business description, security risk factor,s and internal controls reporting), considered here are derived from their textual descriptions and from SEC 10-K filings. I also explore the extent to which simultaneous breaches are correlated with certain firm-level characteristics, such as firm size, revenue, industry type, geographic proximity (based on headquartered states or US regions) and R&D expenditure.

3.3.1 Correlated risk in information security

In a significant work on correlated failures arising as a result of software vulnerabilities shared across organizations, (Chen et al., 2011) propose queuing models for quantifying downtime loss as a function of investment in security technologies, software diversification, and IT resource investments. They further model and analyze the effectiveness of software diversification strategies to deal with correlated failures from different cost-benefit perspectives. (Kunreuther & Heal, 2003) develop game-theoretic models to address the problem of interdependent security where all agents are identical for different real-life scenarios such as airline security, fire protection, vaccinations, and protections against theft and bankruptcy.

Correlated security risks are widely investigated in the domain of cyber-insurance literature as an effective mechanism for minimizing and managing cybersecurity incidents. Businesses must consistently manage adverse events. In the context of cybersecurity, breaches are adverse events involving information security that must be managed, resolved, and prevented in the future. Alongside the rapid growth in Internet, e-commerce, and software capacities, and widespread financial losses following breaches, insurance companies started developing specialized cyber-insurance policies in the late nineteenth and early twentieth centuries (Baer & Parkinson, 2007). Because cyber-attacks and information security breaches often exploit shared vulnerabilities across interconnected networks, interdependent security risks have hindered the development of an uniform cyber-insurance market (R. Anderson & Moore, 2009; Ross Anderson &

Moore, 2007; Böhme & Kataria, 2006). In their empirical analysis of cyber risks in comparison with operational risks, (Biener et al., 2015) pointed out the difficulty of insuring against cyber risks due to the interconnectedness of computer networks and information systems. (Mukhopadhyay et al., 2013) propose security models designed to evaluate the utility of cyber-insurance products based on the concepts of collective risk modeling theory. (Hossack, Pollard, & Zehnwirth, 1999) argue that the financial benefits of cyber-insurance for businesses and organizations far outweigh the costs associated with IS breaches. Analysts of insurance policies and markets in general use collective risk models to calculate premiums charged by the insurer[20] for a given portfolio (Heckman & Meyers, 1983; Meyers & Schenker, 1983). According to (Meyers & Schenker, 1983), collective risk modeling theory attempts to calculate the likelihood that a loss based on an insurance contract exceeds a given amount. The underlying calculations for loss are based on the severity of claims and their count distributions (Meyers & Schenker, 1983). As emphasized in the discussion section, the practical implications of this study include enabling actuaries to incorporate diverse business characteristics during calculations of risk premiums (i.e., aggregate losses) based on the assumption that a breach may result in recalculation (and probably increase) of premiums for other similar firms that share characteristics**.** (Böhme & Kataria, 2006; Böhme & Schwartz, 2010; Öğüt et al., 2011) provides various frameworks for modeling correlated risks in the context of cyber-insurance.

3.3.2 Variable Description

Breach proximity—or, breach relatedness—is the likelihood that two firms or businesses may be breached together (insofar as breaches are reported as related to said firms in a public announcement in our dataset). In the context of this study, a simultaneous breach affecting two related businesses occurs on the same day at each firm. The simplest examples of two firms being breached together are either 1) a denial of service attack or hacking attack with multiple systems breached across more than one organization on the same day and time; or 2) a virus affecting websites of multiple organizations. A third example entails the improper

[20] The algorithm for determining insurance premiums for consumers is based on insurers' calculations of aggregate losses or loss distributions among their client pools.

disposal of consumers' personal records from multiple organizations so that identity thieves, hackers, or criminals may then later misuse those records. As mentioned above, although research exists on correlated security failures, interdependent cybersecurity, and correlated risks from technological and cyber-insurance perspectives, nearly all those studies employ a game-theoretic modeling approach. Alternatively, this dissertation adopts a data science perspective to empirically test the phenomenon of correlated failures and concurrent breaching using concrete industry data. Also, instead of considering technology-based variables like shared vulnerability across software or interconnected computer networks, I explore underlying firm-level antecedents contributing to breach proximity.

The concepts of business similarity and relatedness of businesses have been used as an antecedent in studies on mergers and acquisitions (Shi et al., 2015; L. Wang & Zajac, 2007) and alliance formation (Stuart, 1998). These studies presuppose that businesses that share products, markets, and/or technological space can achieve business synergy easily and hence have higher probability of success when merged or made partners in comparison with dissimilar firms. Historically, business similarities are measured on criteria defined by organizations other than those businesses. For example, the industry classification codes of Standard Industrial Classification (SIC) and North American Industry Classification System (NAICS) are used in the context of business mergers and acquisition (Mathew L. A. Hayward, 2002; L. Wang & Zajac, 2007). Alternatively, business similarity has been understood by counting overlapping patents as a means to measure "technology overlap" (Mowery, Oxley, & Silverman, 1998). Few studies also rely on financial measures, such as those from COMPUSTAT or annual financial reports, to determine business similarity (Altman, 1968; Robins & Wiersema, 1995). Table 3.1 summarizes previous research regarding firm similarity that considered firm characteristics as defined by outside institutions.

Article	Key Variables	Key Findings
(Shi et al., 2015)	Dyadic business proximity based on business descriptions. Dyadic relationships between businesses based on mergers & acquisition, job mobility and investments.	Mergers and acquisitions in the US high tech industry, along with investment and job mobility are significantly correlated with business proximity based on business descriptions.
(L. Wang & Zajac, 2007)	Occurrence of an alliance or an acquisition is used as dependent variable. Business similarity and complementarity based on NAICS codes along with alliances formed in last five years are used as independent variables. Relative size, relative ROA and industry sectors are few of the control variables.	Resource similarity, complimentarity, combined relational capabilities and partner specific knowledge between a pair of firms have an affect on the likelihood of pair of firms forming an alliance or engaging in an acquisition.
(Stuart, 1998)	Formation of alliances in semiconductor industry is used as dependent variable. Positions of firms (crowded vs. prestige) are used as independent variables.	Technological positions in terms of crowding and prestige predicts alliance formation, both at the firm level as well at the dyad level.
(Mowery, Oxley, & Silverman, 1996)	Transfer of technological capabilities based on citation patterns is used as a measure of interfirm knowledge transfer. Forms of alliance formations (equity joint ventures contract-based alliances) are used as dependent variables.	Various hypotheses based on interfirm knowledge transfer in alliances were tested. Authors found technological capabilities of partner firms becoming more divergent in majority of alliances.
(Mowery et al., 1998)	Technological overlap between firms is measured using patent citation data to predict partner selection during inter-firm collaborations.	Partner selection can be predicted from technological overlap. It was also found that alliances influence technological capabilities of firms.
(Gulati, 1998)	This article presents a social network perspective to strategic alliances and emphasizes on the role of precursors, processes and outcomes associated with alliances.	The articles highlight current network research on organizational alliances and provides a road map for future alliance research.
(Lavie, 2007)	This study is aimed as studying the effect of alliance portfolio on firm's market performance for the software industry.	Author found the market performance of focal firm to improve with intensity of competition among partners.

Table 3.1 Previous studies based on business similarity in different contexts

In a recent study on business similarity, (Shi et al., 2015) developed a measure of dyadic business proximity for technology firms based on business descriptions. Specifically, (Shi et al., 2015) defined dyadic relationships in the US technology industry between firms based on mergers and acquisitions, job mobility, and investments, each of which categories are significantly correlated with business relatedness.

Through the office of the Securities and Exchange Commission (SEC), United States federal law requires public firms to report consistently on their financial information by way of quarterly reports (form 10-Q) and annual reports (form 10-K). An annual 10-K report is different from an annual shareholders report and provides a comprehensive overview of the business and its financial condition. Since 10-K financial reports include characteristic details about public firms operating in the US, they can be used to measure business similarity.

(Hoberg & Phillips, 2016) identified related firms based on the business descriptions sections of their 10-K filings. As per SEC policy, the business descriptions section of the 10-K filings includes significant products offerings by businesses, and hence firms offering similar products can be grouped together based on these filings. Analogous to existing external industry classification paradigms for public firms, such as SIC and NAICS, (Hoberg & Phillips, 2016) propose a system of text-based network industry classification based on descriptions from 10-K filings. (Hoberg & Phillips, 2010) found that transactions based on mergers and acquisitions between firms that use similar product descriptions in 10-K filings are more alike than firms for which product descriptions are dissimilar. The research of (Hoberg & Phillips, 2010, 2016) suggests that scholars can appropriately use 10-K filings as a measure of industry classification, as similar firms share similar descriptive content in those forms. No empirical study has yet used 10-K descriptive content as a measure of firm similarity in the context of information security breaches. This study is then unique in that I contend that business proximity can predict that proximate businesses experience higher risk of being breached together.

Public firms often disclose security risk factors associated with their information systems resources in their 10-K public filings. For example, the list of companies in this analysis includes Automatic Data Processing

(ADP), and the security risk factors section from ADP's 10-K filings states that "cybersecurity and privacy breaches may hurt our business, damage our reputation, increase our costs, and cause losses." In addition, one of Twitter's security risk factor sections states, "We are unable to combat spam or other hostile or inappropriate usage on our platform." In a significant work on a firm's security risk factor as stated in its 10-K filing and future disclosure of breach announcement by public firms, the decision tree based model proposed by (T. Wang, Kannan, & Ulmer, 2013) associated disclosure of security risk factors with future breach announcements. My research is similar to that carried out by (T. Wang et al., 2013), in that both studies associate disclosures of security risk factors from 10-K filings with future breach announcements. However, this dissertation is unique because I consider the relationship between security risk factors and a firm's public disclosure of breaches from the social network perspective of firms being breached together. The unit of analysis in this study is not the individual firm but the dyadic relationship between two or more. Thus, I propose that firms that firms with similar security risk factors as disclosed in their public filings are more likely to be the victims of the same breach on the same day (i.e. breaching together) in future in comparison with firms with dissimilar security risk factors.

Internal control reporting by an independent third-party auditor is a mandatory form of disclosure required for firms in their SEC 10-filings. Specifically, sections 302 and 404 of the Sarbanes-Oxley Act of 2002 require public companies to have their financial reporting audited by an independent auditor whose report is included in the 10-K filing. The main purpose of reporting on internal control is the establishment, maintenance, and evaluation of financial reporting of the public organization by an independent auditor so that no discrepancies or misrepresentation of financial information occur in filings (Ge & McVay, 2005). The accounting literature on disclosures shows that one of the aims of reporting internal controls is to disclose material weakness of the public firm(Ge & McVay, 2005). The Public Company Accounting Oversight Board (PCAOB), defines such weakness as "a deficiency, or a combination of deficiencies, in internal control over financial reporting, such that there is a reasonable possibility that a material misstatement of the company's annual or financial statements will not be prevented or detected on a timely basis."[21] (Doyle, Ge, & McVay, 2007) analyzed some of the determinants of weaknesses in internal controls for public firms. Less severe than material weakness is a significant deficiency, which is another measure of internal vulnerability and which, if present, should be reported mandatorily in the internal control reporting for public firms. Most firms in our dataset reported no significant deficiencies. However, four firms did report a significant deficiency in the year prior to breach, and two firms reported a significant deficiency in the same year as breach. None reported material weakness. Because these vulnerability measures were so limited in my dataset, I was unable to use similarities of firms based on material weakness or significant deficiencies as reported in 10-K. Therefore, I rely primarily on textual analysis of the internal control report to measure firms' relatedness. Firms that are similar in the internal control reporting structure are more likely to be breached together than those with dissimilar reports.

Internal control for a public firm is evaluated based on what is known as the Committee of Sponsoring Organizations of the Treadway Commission or the COSO framework,[22] an industry standard determining the effectiveness of firms' operations procedures, the rigidity of their compliance with laws and regulations, and the reliability of their financial reporting.[23] As per the COSO framework, an effective internal control system within an organization consists of five components: control environment, risk assessment, control activities, information and communication, and monitoring. The control activities component encompasses a sub-component for security, which examines a firm's application and network.[24] Recently industry reports assess the COSO framework for internal control in the developing landscape of cyber risk and cyber-attacks.[25]

---

[21] https://pcaobus.org/Standards/Auditing/Pages/Auditing_Standard_5_Appendix_A.aspx
[22] https://www.coso.org/Pages/default.aspx
[23] Reliability of financial reporting is evident in a firm's compliance with SOX sections 302 and 404.
[24] https://info.knowledgeleader.com/bid/161685/what-are-the-five-components-of-the-coso-framework
[25] https://www.coso.org/documents/COSO%20in%20the%20Cyber%20Age_FULL_r11.pdf , https://www.iasa.org/iasadocs/Chapters/Northeastern/Presentations1115/5.1%20IASA%20CT%20Event%20-%20Changing%20the%20Cyber%20Game.pdf

The argument of using internal control reporting in this study on breaches is also supported by the following example of IT related weakness as disclosed in the internal control section of 10-K. This example is linked with a major cyber breach incident, as disclosed by Equifax in its 2017 filing:

"As discussed in Note 5 of the Notes to the Consolidated Financial Statements in this Form 10-Q, on September 7, 2017, we announced a cybersecurity incident. Our review of the circumstances and resulting impact on our internal controls over financial reporting (ICFR) identified two significant deficiencies in our IT General Controls environment, at this point in time. As part of the Company's overall plan to address the cybersecurity incident, actions have already been and are being taken in the fourth quarter of 2017 to remediate these significant deficiencies."[26]

This dissertation chapter makes clear that internal control reporting does contain information systems, cybersecurity, and cyber-attacks components in the form of IT disclosure, which makes this variable relevant as a measure of business similarity in the context of security breaches and as an antecedent to firms being breached together.

In this study, I have also used similarity of firms based on their absorptive capacity. I posit that firms with similar absorptive capacity have similar capabilities to acquire, develop, and use knowledge from external and internal resources. First proposed by (Cohen & Levinthal, 1990), absorptive capacity, as an organizational concept, is defined as "a firm's capability to recognize the value of new, external information, assimilate it, and apply it to commercial ends." Historically, absorptive capacity has been used to explain various organizational level phenomena, such as inter-organizational learning (Peter J. Lane & Michael Lubatkin, 1998) and knowledge management and transfer (Frans A. J. Van den Bosch, Henk W. Volberda, & Michiel de Boer, n.d.; Tsai, 2001; Zahra & George, 2002). Absorptive capacity has also been used in the information systems domain. For example, (Roberts, Galluch, Dinger, & Grover, 2012) provide a literature review of absorptive capacity's use in IS. Consistent with previous research, where a firm's

[26] https://www.sec.gov/Archives/edgar/data/33185/000003318517000032/efx10q20170930.htm#sCE0A5681DCD352C6AC71478407A9A6E3

R&D investments are traditionally used to measure its absorptive capacity (Cohen & Levinthal, 1990; Rachel Griffith, Stephen Redding, & John Van Reenen, 2003), I have used it as a similarity measure for firms and argue that firms with similar R&D intensity (Deeds, 2001) will have similar absorptive capacity. As such, firms with similar levels of absorptive capacity will have similar knowledge bases (and security measures) against cyber threats, and they are thus more likely to be breached together than firms with dissimilar R&D intensity. Following prior empirical studies, R&D intensity was measured as the ratio of R&D expenditure and sales.

Traditionally, firm or business size has been measured in terms of revenue, assets, market share, and number of employees. The size of the firm had been used in previous management and IS studies in different contexts. For example, firm size has been viewed as an enabler of innovation and R&D efforts of firms (Damanpour, 1992; Rogers, 2004; Shefer & Frenkel, 2005). Other studies from the information system literature have utilized firm size to study outsourcing potential (Grover, Cheon, & Teng, 1994; Loh & Venkatraman, 1992) or IT innovation (Teng, Grover, & Guttler, 2002).

The literature also suggests that firms of similar sizes (regarding revenue or number of employees) and industry classification have similar organizational capabilities, including technological ones (Wignaraja, 2002). In this study, I used the first two digits of the SIC codes for industry classification to group similar firms, as suggested in (T. Wang et al., 2013). From the perspectives of resource-based and dynamic capabilities of technological firms, IT capability (Bharadwaj, 2000), software and computer capability (Cron & Sobol, 1983), and project management capabilities (Ethiraj, Kale, Krishnan, & Singh, 2005) were found to have a positive relationship with firm performance. Because these studies involving technology-based firms controlled effectively for different firm sizes, they suggest that businesses similar in size may have similar technological capabilities. Furthermore, such technologically similar firms with similar IT capabilities will require software with similar functionalities to be installed over their telecommunication networks to carry out their day-to-day operations. Thus, these firms may share software stacks, and that sharing scenario could result in shared software vulnerabilities (Chen et al., 2011). For instance, a big business with thousands of employees working in the health insurance industry may not only have similar operationalization of their business process routines as other health insurance firms, but they may also have similar software configurations to run business processes. Additionally, a recent industry report suggests correlation between firm size and IT spending for small and medium US businesses.[27] This correlation could imply that similar sized firms have similar IT capabilities and may have similar software stacks installed on their information systems. These proximities are also evident from observations of firm dyads in my dataset. For instance, Citi Group, Inc. and Bank of America, or HP and Symantec, which are large firms belonging to the same industrial sector, have similar operations procedures and technological capabilities.

Geographic or spatial proximity may be defined as the relatedness of businesses in terms of geographic closeness. Geographic and spatial proximities have been used in the past as measures of business relatedness in the context of mergers and acquisitions (Shi et al., 2015), innovation (Ben Letaifa & Rabeau, 2013; Boschma, 2005), and economic advantages (Ellison & Glaeser, 1997). Most previous studies from management and IS examined the benefits of geographic proximity or being co-located with other business. In contrast, my study explores the potential for disadvantages (especially, breaches) of geographic or spatial proximity with other businesses.

I hypothesize that it is more likely that firms which are co-located in terms of headquartered state or US regions (Midwest, Northeast, South, and West, as per US census bureau[28]) will be victims of the same information security breach compared with firms headquartered in different states or regions. My hypothesis is reasonable given that computer science studies have investigated the effects of geographically proximal network systems on correlated failures and network reliability of fiber optic networks (Neumayer & Modiano, 2010) and power grid systems (Bernstein, Bienstock, Hay, Uzunoglu, & Zussman, 2011).

[27] https://www.comptia.org/resources/it-industry-trends-analysis
[28] https://www.census.gov/geo/reference/webatlas/regions.html

The relationship between geographic proximity of firms and security breach relatedness can also be argued from the viewpoint of a theory on regional security known as regional security complexity theory (Buzan, 2008; Buzan & Waever, 2003). Although it was originally developed for international relations, this regional security theory includes the basic concept of regional security complexes and can also be employed in the context of information security. The concept of regional security complexes intuits that security concerns are usually concentrated in geographical regions or regional sub-systems intertwined with other sub-systems through a web of security interdependence (Buzan, 2008). According to (Buzan, 2008), these regional concerns, such as military or political threats, do not travel over long distances and are often associated with geographical proximity. Thus, they result in the formation of regionally-based clusters known as regional security complexes. Inherently, security interdependence between states within a regional security complex is more entwined than between states both inside and outside the regional security complex (Buzan, 2008). We hypothesize that if a firm shares similar traits with others, in addition to geographic proximity, then their likelihoods of being breached together will increase. Table 3.2 provides examples of prior studies that have used the firm-level attributes included in our study.

Independent variable	Explanation (data source)	Previous literature
Firm descriptions from 10-K filings	Description section in 10K includes information about major product offerings (SEC EDGAR system).	(Hoberg & Phillips, 2010, 2016)
Risk factors and security risk factors from 10-K	Description of the disclosed security risk factors from the 10K filed in the previous year as when breach happened (SEC EDGAR system).	(T. Wang et al., 2013)
Internal control report from 10-K	Description of the internal control in the form of a report by an independent audit company (SEC EDGAR system).	(Ashbaugh-skaife & Lafond, 2006; Doyle et al., 2007)
R&D intensity (ratio of R&D expenditure to sales)	R&D expenditure is historically being used as a proxy for a firm's absorptive capacity and organizational learning (COMPUSTAT).	(Bardhan, Krishnan, & Lin, 2013; Cohen & Levinthal, 1990; Roberts et al., 2012)
Headquartered state	As a measure of geographic proximity.	(Shi et al., 2015)
Headquartered US region	As a measure of geographic proximity (US geographic regions as per US census).	(Leidner & Kayworth, 2006)
Industry type	Type of industry indicated by SIC code for industry classification (COMPUSTAT[29]).	(L. Wang & Zajac, 2007)
Firm size (# of employees)	COMPUSTAT	(Altman, 1968; Robins & Wiersema, 1995)
Firm revenue	COMPUSTAT	(Altman, 1968; Robins & Wiersema, 1995)

Table 3.2. Independent variables with data source and literature

[29] https://wrds-www.wharton.upenn.edu/pages/

3.4 Data Collection

The information security breach data for the firms that have announced breaches together on the same day from 2008–2017 was collected from the Privacy Rights Clearinghouse dataset. Specifically, the data include the date the breach was made public, victim firm(s), location(s) of breach, type of breach, and a short description of events.[30] The breach dataset from the Privacy Rights Clearinghouse is a popular and publicly-available source of breach information for IS researchers (Avery & Ranganathan, 2016; Kwon & Johnson, 2015; Sen & Borle, 2015). As noted, I collected breach data for this study in July 2017.

The short descriptions of events in the breach dataset reveal that some of the firms in the dataset share client-provider relationships (for example, ADP providing payroll services to US Airways), collaborative/partnership relationships (such as HP and Symantec, with the latter acting as IS provider to the former), or competitive relationships (such as Bank of America and Citi Group, Inc.). As argued by (Chen et al., 2011), for partner firms, use of homogenous software offers many positive network effects, such as increased compatibility and interoperability.

I collected 10-K filings on breached firms from the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system for the years in which breaches occurred. I also collected 10-K filings from the previous year to observe independent variables, including security risk factors and internal controls reporting. As mentioned above, 10-K filings contain firm description, security risk factors, and internal control reporting, all of which can be analyzed textually to measure firm relatedness. Characteristics including firm size (number of employees), revenue, R&D expenditure, and industry type (SIC code) were collected from COMPUSTAT. The resulting network consists of 33 publicly listed firms, which have been breached together from 2008–2017. Headquartered state or US region are collected from public sources such as the firm's public website or via Google searches.

[30] https://www.privacyrights.org/data-breaches

While analyzing data, I experimented with an independent variable of firm similarity based on the independent auditing firm.[31] I hypothesized that firms with the same independent auditor might have been breached together. Data about the auditing firm in the breached year is collected from the WRDS's audit analytics database. However, I found this variable insignificant.

For the same years in which breaches occurred, I collected data, including independent variables, firms' 10-K descriptions, SIC codes for industry classification, firm revenue from COMPUSTAT, and number of employees from COMPUSTAT. Consistent with (T. Wang et al., 2013), security risk factors were collected from the annual 10-K filings from the year prior to the occurrence of the breach. Internal control reporting by independent auditor firms and management's reports on internal control were also collected from the previous year's 10-K. The variable representing the absorptive capacity of the firm in the form of research intensity, which is the ratio of R&D expenditure to sales and returns on assets, were also collected for the year prior to the breached year. I collected these variables (such as security risks, internal control, and R&D intensity) from the previous year because firms with similar organizational characteristics based on these variables run a higher risk of being subsequently breached together. On the other hand, I collected from the same year the variables, firm size (number of employees) and revenue, because they motivate current breaches rather than subsequent ones. For example, our breaches dataset includes insider breaches, theft, and unintended disclosure, which likely depend more on current employees (when the breach occurred) than on previous years' employees.

A complete list of firms' SIC codes, headquartered states, and headquartered regions is located in Appendix B, Table B.1. Table 3.3 shows the distribution of firms by 2-digit SIC codes for firms breached together in our dataset.

---

[31] The five auditing firms are Deloitte & Touche (n=7), Ernst & Young (n=9), KPMG (n=4), Moss Adams (n=1), and PricewaterhouseCoopers (n=12).

2-digit SIC code description	Number of firms
73 - Business Services	12
48 – Communications	1
60 - Depository Institutions	2
58 - Eating and Drinking Places	1
36 - Electronic & Other Electrical Equipment & Components	4
34 - Fabricated Metal Products	1
35 - Industrial and Commercial Machinery and Computer Equipment	1
64 - Insurance Agents, Brokers and Service	1
63 - Insurance Carriers	3
61 - Nondepository Credit Institutions	2
29 - Petroleum Refining and Related Industries	2
27 - Printing, Publishing and Allied Industries	1
45 - Transportation by Air	1
51 - Wholesale Trade-Nondurable Goods	1

Table 3.3. Industry distribution of firms by 2-digit SIC codes

The summary statistics for quantitative independent variables representing firm characteristics used in this study are shown in table 3.4.

Independent Quantitative Variable	Minimum	1st quartile	Median	Mean	3rd quartile	Max	Standard deviation	Variance
Revenue (in millions)	3.966	5299.0	64851.7	105790.0	105790.0	470171.0	92815.77	8614766259
R&D intensity (R&D to sales)	0.00	0.00	0.0040	0.0711	0.0980	0.4480	0.113021	0.012774
Number of employees (in 1000s)	0.006	12.200	49.400	112.916	167.900	433.362	135.603	18388.08
Return on Assets	-0.0644	0.05000	0.11000	0.1144697	0.169000	0.286000	0.093917	0.008820

Table 3.4. Summary statistics for quantitative independent variables

As mentioned previously, risk factors and security risk factors for the firms were collected from 10-K filings. I was able to identify 245 security risk factors across 33 firms with a mean of 7.42 per firm and standard deviation of 7.41. Note that these security risk factors are not unique. The five topics that I extracted through topic modeling and their associated keywords are summarized in table 3.5.

Topic #	Keywords	Underlying theme
0	Business, systems, information security, data, customers, operations, loss, financial service, reputation, failure, parties, result, adverse	Business data systems, information security, and failures
1	Products, a software company, proprietary, adversely, system failures, source, problems, license, bugs, sales, operating, foreign, errors, results, customers, open vulnerabilities, hardware	Software products, system failures, bugs and vulnerabilities
2	Regulations, information privacy, providers, requirements, health vendors, federal, state, fail, comply, relating, standards, healthcare and/or phi	Information and healthcare standards, compliance and privacy
3	Users, services, products, data privacy, regulations, laws, access, operating practices, online software infrastructure, content, user, protection	User data privacy, protection, and online content
4	firm risk, operational, financial, clients, enterprise management, confidential, infrastructure risks, control events, transactions, process, losses, reputational, practices	Firm operational, infrastructure and financial risk

Table 3.5. Topic-based important keywords for security risk factors

3.5 Research Methodology

3.5.1 Topic Modeling

Topic modeling is a text mining technique that identifies latent topics or themes from a set of structured or unstructured text documents. More formally, topic models are algorithms that discover underlying themes in a large collection of documents (Blei, 2012). Using this statistical technique, one can identify common topics in a set of documents and organize those documents by theme. One of topic modeling’s advantages is that the methodology does not require pre-labeling of documents to reveal themes. Therefore and the technique constitutes an unsupervised machine learning approach to document classification and retrieval (Blei, 2012). The collection of documents in this study is the set of the 10-K filings and corresponding business descriptions.

The origins of topic modeling algorithms can be traced back to the development of technique of Latent Semantic Analysis (LSA) or Latent Semantic Indexing (LSI), a method for automatic indexing of documents primarily developed for information retrieval (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Wiemer-Hastings, Wiemer-Hastings, & Graesser, 2004). LSA was proposed with the purpose of retrieving documents based on the conceptual topic or theme or meaning of the document. Hence, LSA was initially developed to solve the problem of matching search terms with terms in a collection of documents and to retrieve relevant documents as per user queries. Synonymy (different words with similar meanings) and polysemy (the same word with multiple meanings) are the two main problems it tries to solve (Deerwester, Dumais, Furnas, Landauer, & Harshman, n.d.; Hofmann, 1999). LSA approaches the problem of associating terms with documents as a statistical problem and uses the technique of Singular Value Decomposition (SVD). SVD is a matrix decomposition technique similar to statistical techniques, such as eigenvector decomposition and factor analysis based on linear algebra, for analyzing matrices of terms (as term-document matrices or co-occurrence tables) for the documents corpora (Deerwester et al., n.d.; Golub & Reinsch, 1970).

A probabilistic approach to LSA with sound statistical foundation was developed later and is known as probabilistic latent semantic analysis. It provides more efficient results compared to its non-probabilistic counterpart (Hofmann, 1999). Probabilistic latent semantic analysis is based on the latent variable statistical model known as the aspect model (Hofmann, Puzicha, & Jordan, 1999). Latent Dirichlet Allocation (LDA) is a generative probabilistic model for topic modeling where each topic is characterized by a distribution over words with each word belonging to a vocabulary vector from the corpus, and each document is represented as random mixtures (based on the mixture model) over underlying themes or topics (Blei, 2012; Blei, Ng, & Jordan, 2003).

In LDA, a topic is the underlying latent or hidden variable, defined by distribution over a fixed vocabulary, and is assumed to be generated before the generation of any data. One of the underlying assumptions of LDA is that the number of topics are fixed and need to be assigned by the researcher (Blei, 2012). With LDA, each topic is represented to varying degrees in different documents. That is, each document carries different proportions or probability for every topic. Also, each word in each document is chosen from one of the topics based on the Dirichlet distribution over the topics. The LDA algorithm can be summarized in the following steps:

Step 1: Assign number of topics or underlying latent themes. One can assign a different number of topics to determine if LDA can reveal meaningful word groups from the corpus as topical themes.

Step 2: LDA will assign a temporary topic to every word in the corpus. This temporary assignment is based on the Dirichlet distribution of topics and is updated iteratively in the next step. Note that the number of iterations needs to be specified by the researcher.

Step 3: LDA will iterate through each word in every document and reassign topics to these words based on a word’s prevalence across different topics and on a topic’s prevalence across different documents. For each word and its document, the likelihood of a topic choice depends on two factors: 1) likelihood of a certain topic for a certain document; and 2) likelihood of a certain word for a certain topic (Grus, 2015).

The graphical model for LDA (Blei, 2012) is shown in Figure 3.1.

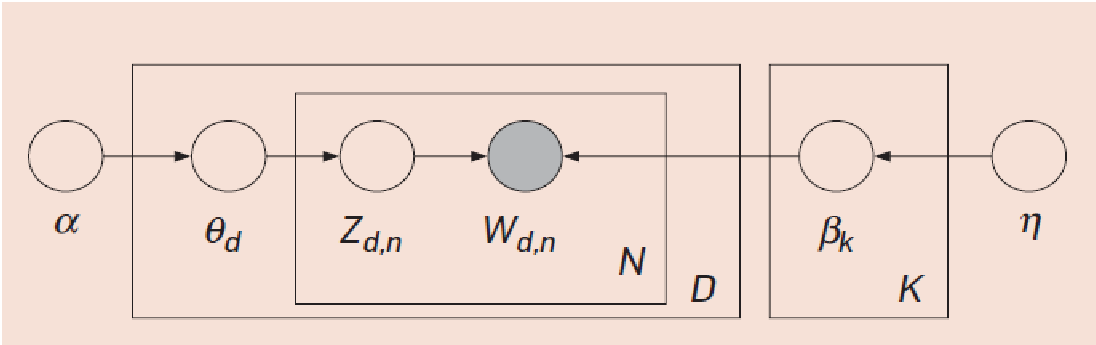


Figure 3.1. A graphical model for LDA (Blei, 2012)

Where K = total number of topics

β_K = distribution of a topic over the vocabulary or the distribution of a topic over the words

D = total number of documents

θ_d = topic proportions for d^{th} document

N = total number of words in a document

$Z_{d,n}$ = topic assignment for n^{th} word in document d

$W_{d,n}$ = n^{th} word in document d

Note that the topic assignment $Z_{d,n}$ depends on θ_d and $W_{d,n}$ depends on $Z_{d,n}$, and $\beta_{1:K}$ for each topic K (Blei, 2012).

LSA, its probabilistic variant, and LDA all rely on the "bag-of-words" assumption according to which the order of occurrence of words in a document does not matter. The order of documents is also irrelevant. (Blei et al., 2003) posit that this assumption is too stringent for practical text analysis and argue that n-grams are more appropriate than unigrams for most practical contexts. LDA is also based on the assumption of exchangeability (D. J. Aldous, 1985) and de Finetti's theorem on exchangeability (D. Heath & Sudderth, 1976). According to de Finetti's theorem on exchangeability, any collection of exchangeable random variables can be represented as a mixture distribution. Similarly, (Blei et al., 2003) consider mixture models

(Figueiredo & Jain, 2002; McLachlan & Peel, 2004) for capturing the exchangeability of words within document and documents within the corpus. Another assumption of LDA is that the number of topics are fixed and need to assigned by the researcher (Blei, 2012).

In this study, I use topic modeling based on LDA to assess the business relatedness of firms that have been breached together in the past 10 years (2008–2017). The application of LDA in this research is similar to applications in (Shi et al., 2015) and (Hoberg & Phillips, 2016).

I used the MALLET (i.e., Machine Learning for Language Toolkit) implementation of the Latent Dirichlet Analysis (LDA) technique to perform topic modeling over our data set of firm descriptions. One of the MALLET outputs provides the probability loadings of each of the individual firm descriptions on each topic. This output can then be utilized to create a cosine similarity matrix for the firm descriptions based on the argument that firms with similar probability distributions across topics would be more like each other than those with dissimilar probability loadings for those topics. The cosine similarity matrix we got from this step can be used as an input to the next step of statistical analysis using the Quadratic Assignment Procedure (QAP).

3.5.2 Quadratic Assignment Procedure (QAP)

Data in most studies involving social networks consist of observations that correspond to a pair of entities (i.e., a dyad) rather than individual entities themselves. Examples include co-authorship networks, citation networks, friendship networks for individuals, a network of import/export relationships between countries, or ties between board members who serve on the boards of the same firm(s). Statistical analyses of such networks entail analysis of dyadic relationships between these social network entities. Such dyadic relationships are dependent on each other. For example, in a simple Facebook network, if A is a friend of B and B is a friend of C, then the likelihood of A becoming a friend of C depends on the existing friendship relationship between A with B and B with C. More formally, the likelihood of formation of a tie in a social network between the network members depends upon their existing attributes along with the structure of the existing network (Hunter, Handcock, Butts, Goodreau, & Morris, 2008b).

The unit of analysis or the outcome of our study is the tie or the relationship between firms rather than the firms themselves. Due to the dyadic nature of ties between firms, the unit of analysis (which is a tie or a relationship) is dependent on both components of the tie. Thus, the regular techniques of linear regression cannot be used. Most studies involving social networks suffer from this inherent challenge. The standard statistical techniques of linear regression cannot be used due to auto-correlation between dyads. Standard regression procedures will result in biased results and incorrect standard errors.

To approach analyses of such data sets and to obtain correct estimates for the standard errors, researchers have used sampling-based approaches such as Quadratic Assignment Procedure (QAP) (Krackhardt, 1992; Simpson, 2001) and Exponential Random Graph Models (ERGM) (Goodreau, 2007; Goodreau, Kitts, & Morris, 2009; Harris, 2013b; Hunter et al., 2008b). In this paper, I employ the network-based statistical procedure of QAP to find the correlation between the business proximity measure derived from firm descriptions and information security breaches.

The non-parametric statistical technique of QAP has been found to produce relatively unbiased results for social network analysis (Krackhardt, 1988). QAP can be used to perform social network analysis for analyzing dyadic data sets where each dyad represents a relationship between two or more social entities. Some of the questions that can be answered using the QAP technique include: 1) Are authors with similar characteristics, such as gender or affiliation, more likely to collaborate compared with authors with dissimilar individual characteristics? 2) Do firms with similar board member composition or some other form of business similarity tend to perform similarly in the stock market? and 3) Are Facebook users more likely to form friendships with other Facebook users who have similar characteristics (for instance, similar gender, demographics, and education)?

The QAP technique works by permuting a graph in such a way that the graph's underlying network structure remains the same, but the rows and columns (i.e., nodes or actors or network entities) are assigned different network positions with each permutation (Prell, 2012). This process of permutation can be repeated multiple times and the correlation coefficients for the dependent and independent matrices are recalculated at every iteration and thus result in a distribution of permuted samples. Finally, the proportion of times the results (in the form of correlation coefficients) from theses permuted matrices are same (or different) from original correlation coefficients can lead us to statistically insignificant (or significant) results. The working of QAP algorithm is summarized as below:

Step 1: Calculate the correlation coefficient for the dependent and independent matrices. Since this coefficient and corresponding p-value(s) with standard errors are incorrect due to auto-correlation carry out step 2.

Step 2: Permute the dependent matrix such that the network structure remains the same, but rows and columns are assigned differently. Then recalculate the correlation coefficients for the dependent and independent matrix (or matrices).

Step 3: Repeat this process of permutations thousands of times with each iteration recalculating the correlation coefficients resulting in sampling distribution. The purpose of the sampling distribution is to take into account any correlation between individual nodes (Simpson, 2001).

Step 4: Hypothesis test whether any correlation exists between the dependent and independent matrix (or matrices). The corresponding significance level is carried out as follows: if the results from original dataset rarely appears in sampling distribution, one can reject the null hypothesis and state that the dependent and independent matrix (or matrices) carries a correlation which is statistically significant (Prell, 2012; Simpson, 2001).

The steps used in this research are summarized in figure 3.2.

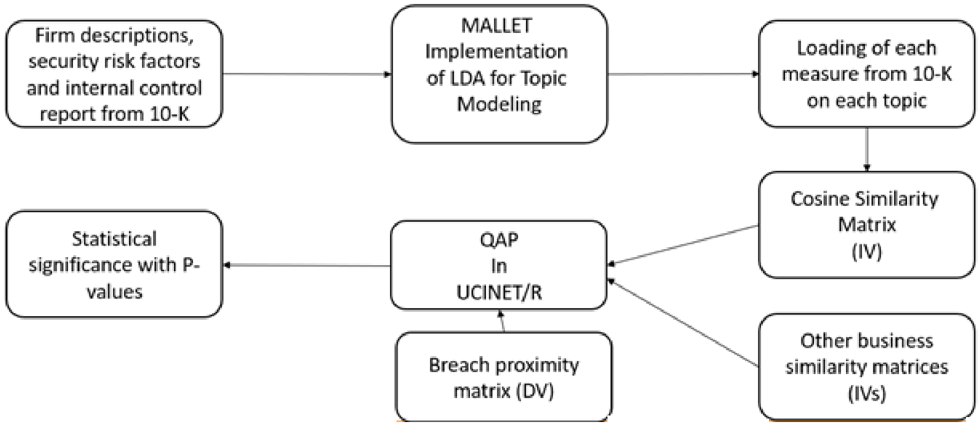


Figure 3.2. Steps of research methodology

3.6 Analysis and Results

As stated previously, QAP, a popular statistical technique employed by social network analysis researchers, was used to test for correlation and significance of the association between the firms that are breached together and between various business similarity measures. In statistical parlance, the model comprises a dependent matrix of firms that have been breached together in the past 10 years. The independent matrices represent similarity of firms based on different firm characteristics such as industry type and headquartered state/region. A matrix also exists for cosine similarity based on SEC 10-K filings (for firm descriptions, security risk factors, and internal control reporting). In this research, I used the US state and the region in which the firm is headquartered as a measure of geographic proximity.

The dependent matrix is a binary matrix consisting of values 0 (not breached together) or 1 (breached together). Except for the independent variables derived from the text of SEC 10-K filings, all the independent matrices were constructed in the same way, with a 1 indicating firm similarity and 0 otherwise. The values in independent cosine similarity matrix based on 10-K filings are all continuous and less than 1 with diagonal values ignored. To carry out the hypotheses tests, these dependent and independent matrices are given as inputs to the QAP procedure of a well-known software for social network analysis called UCINET.[32] Basic social network analysis is performed using various statistical packages for social network analysis in R statistical computing platform[33] using RStudio.[34]

The breach network consists of 33 public firms affected by the same breach on the same day. These businesses are listed in Appendix B, Table B.1. Also, the adjacency list of our breach network, along with the firm names and the date on which the breach was made public, is shown in Table B.2 of Appendix B.

---

[32] https://sites.google.com/site/ucinetsoftware/home

[33] R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[34] RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/

The breach network is shown in Figure 3.3, and some of the basic network characteristics are shown in Table 3.5. The degree centrality of all nodes is the same (0.03125), except for the nodes representing Automatic Data Processing (0.1875), US Airways (.0625) and McDonald's (0.625). Similarly, the betweenness centrality of every node is zero except for these three nodes: Automatic Data Processing (0.040322581), US Airways (.01296774) and McDonald's (.002016129).

# of Vertices or network size	33
# of edges	20
Network density	0.03787879
Network degree centralization	0.1592742
Network betweenness centralization	0.03988155
Network closeness centralization	0

Table 3.6. Basic network characteristics for breach network of public firms

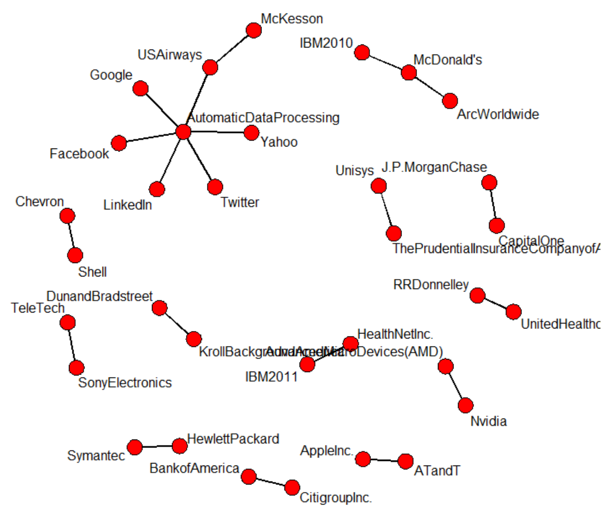


Figure 3.3. Breach network for 33 public firms

Note that there are two separate nodes for IBM because IBM experienced two separate breaches in years 2010 and 2011. I used the corresponding SEC 10-K filings from IBM for both the years for the respective breaches in our analysis.

Within UCINET, I used the multiple regression variant of QAP known as MR-QAP, where the dependent matrix is the breach network for the 33 public businesses and independent matrices are cosine similarity matrices from 10-K filings, geographic proximity matrix (0 for firms headquartered in different states and 1 for those headquartered in the same state), matrix of business similarity based on 2-digit SIC codes from COMPUSTAT (again 0 for firms with different SIC codes and 1 for exactly same SIC code), and matrix representing similar firms based on number of employees (represented as categorical variable as obtained from COMPUSTAT).

The results of MR-QAP analysis for four models with their respective p-values are shown in Table 3.7.

Dependent variable: breach_proximity_matrix				
	Model (1)	Model (2)	Model (3)	Model (4)
cosine_similarity_ matrix_preprocessed_ 10K_description	0.02488** (0.01272)	0.02529** (0.01263)	0.0249** (0.01284)	0.02492** (0.01241)
security_risk_factors_cosine_similarity_matrix				0.03583** (0.02076)
risk_factors_cosine_similarity_matrix	0.0326 (0.02873)	0.03154(0.02871)	0.02954 (0.02911)	
similarity_matrix_based_on_ auditor_internal_control_report	-0.01281 (0.01832)	-0.01197(0.01848)	-0.01162 (0.01814)	-0.01349 (0.01847)
similarity_matrix_SIC_2digit	0.01962* (0.01237)	0.01827* (0.01218)	0.01881* (0.01227)	0.02087* (0.01213)
geographic_proximity_ matrix_HQ_state		-0.01057 (0.01071)	-0.01122 (0.01036)	
similarity_matrix_R&D_to_sales	0.01317* (0.00882)	0.01335* (0.00892)	0.01344* (0.00876)	0.01267* (0.00867)
number_of_employees_similarity_matrix	-0.01116 (0.00877)	-0.01155* (0.00866)	-0.01172* (0.00861)	-0.01125* (0.00858)
similarity_based_on_mgmt_internal_control_report	-0.00196 (0.01154)	-0.00448 (0.01107)		-0.0046(0.01157)
geographic_proximity_matrix_HQ_region	-0.0117 (0.00956)			-0.01182 (0.00967)

Table 3.7. Results from MR-QAP analysis – Coefficients with standard errors in brackets

(p<0.1 *, p <0.05 **, p<0.01 ***)

Interpretation as logit or linear probability model when the dependent variable is binary—that is, firms breached together have a link (1); otherwise no link (0)—is done as in the following example, for Model 1:[35]

1. If two firms are similar in terms of their business description from the 10-K, this increases the probability of being breached together by 0.02488 or 2.488%.
2. If two firms are similar in terms of their security risk factors from the 10-K, this increases the probability of being breached together by 0.03583 or 3.583% in the following year.
3. If two firms are similar in terms of their 2-digit SIC code for industry classification, this increases the probability of being breached together by 0.01962 or 1.962%.
4. If two firms are similar in terms of their R&D intensity (i.e. ratio of R&D spending to sales), this increases the probability of being breached together by 0.01317 or 1.317 % in the following year.

The preprocessed textual description for firms from 10-K is obtained by converting all text from firm description to lowercase and removing numbers, punctuations, and stop words. These pre-processed firm descriptions are given as input to Python's topic modeling package (MALLET), followed by construction of cosine similarity matrix, which is finally given as input to UCINET's MR-QAP procedure along with other firm-level characteristics as independent matrices (or variables).

The MR-QAP results for the four models are in Figure B.3 of Appendix B. The four-firm level independent variables consistently having a statistically significant relationship with breaching together in all MR-QAP models are: 1) similarity of firms based on their business descriptions from 10-K; 2) similarity based on security risk factors; 3) similarity based on industry type; and 4) similarity of firms based on their research intensity (or similarity based on their absorptive capacity). The independent variable of firm size (number

[35] http://faculty.ucr.edu/~hanneman/nettext/C18_Statistics.html

of employees) is also significant in the three models. Also, similarity based on geographic proximity (US state and region) and similarity based on security risk factors are close to statistical significance at 90% CI.

The QAP correlations for the independent variables used in the Model 4 are shown in Table 3.8 and the corresponding p-values for these variable correlations in Table 3.9.

	33 nodes cosine matrix on preprocessed firm descriptions from 10K	33 nodes security_risk_factors cosine matrix	33 nodes report of independent auditor on IC matrix	33 nodes management_report_on_IC matrix	33 nodes SIC matrix 2 digit	33 nodes R&D to sales matrix previous year median cutoff	33 nodes number of employees from WRDS categorical matrix median cutoff	33 nodes HQ_region matrix
33 nodes cosine matrix on preprocessed firm descriptions from 10K	1	0.225	-0.037	0.111	0.282	0.263	0.071	0.062
33 nodes security_risk_factors cosine matrix	0.225	1	0.019	0.254	0.064	0.157	0.058	0.057
33 nodes report of independent auditor on IC matrix	-0.037	0.019	1	-0.034	-0.103	0.017	-0.026	-0.034
33 nodes management_report_on_IC matrix	0.111	0.254	-0.034	1	-0.045	0.028	0.099	0.295
33 nodes SIC matrix 2 digit	0.282	0.064	-0.103	-0.045	1	0.184	0.077	0.122
33 nodes R&D to sales matrix previous year median cutoff	0.263	0.157	0.017	0.028	0.184	1	0.082	0.098
33 nodes number of employees from WRDS categorical matrix median cutoff	0.071	0.058	-0.026	0.099	0.077	0.082	1	0.276
33 nodes HQ_region matrix	0.062	0.057	-0.034	0.295	0.122	0.098	0.276	1

Table 3.8. QAP correlations between independent variables

	33 nodes cosine matrix on preprocessed firm descriptions from 10K	33 nodes security_risk_factors cosine matrix	33 nodes report of independent auditor on IC matrix	33 nodes management_report_on_IC matrix	33 nodes SIC matrix 2 digit	33 nodes R&D to sales matrix previous year median cutoff	33 nodes number of employees from WRDS categorical matrix median cutoff	33 nodes HQ_region matrix
33 nodes cosine matrix on preprocessed firm descriptions from 10K	0.000 0	.0030***	.347 0	.0310**	.001 0***	.000 0	.059 0*	0.119
33 nodes security_risk_factors cosine matrix	0.003***	0	0.393	0	0.225	0	0.055*	0.162
33 nodes report of independent auditor on IC matrix	0.347	0.393	0	0.311	0.112	0.186	0.074*	0.324
33 nodes management_report_on_IC matrix	0.031**	0	0.311	0	0.204	0.162	0.036**	0
33 nodes SIC matrix 2 digit	0.001***	0.225	0.112	0.204	0	0.003***	0.047**	0.026**
33 nodes R&D to sales matrix previous year median cutoff	0	0	0.186	0.162	0.003***	0	0.084*	0.043**
33 nodes number of employees from WRDS categorical matrix median cutoff	0.059*	0.055*	0.074*	0.036**	0.047**	0.084*	0	0.001***
33 nodes HQ_region matrix	0.119	0.162	0.324	0	0.026**	0.043**	0.001***	0

Table 3.9. QAP p-values for correlations between variables between independent variables

($p<0.1$ *, $p<0.05$ **, $p<0.01$ ***)

Although many of the independent variables in the MR-QAP analysis are correlated with each other, the analysis carried out in this study using UCINET's version 6 MR-QAP analysis is based on a permutation method called double semi-partialing or DSP (Dekker, Krackhardt, & Snijders, 2007), which is an extension to the previous permutation technique of semi-partialing (Dekker, Krackhardt, & Snijders, 2003). The techniques of partialing and semi-partialing were developed for MR-QAP analysis to remove biased estimates under multicollinearity. The technique of semi-partialing is proposed by (Dekker et al., 2007) and incorporated with UCINET's MR-QAP analysis (TOOLS → TESTING HYPOTHESIS → DYADIC (QAP) → MR-QAP LINEAR REGRESSION → DOUBLE DEKKER SEMI-PARTIALING MRQAP). The aim of carrying out permutations of network based on semi-partialing technique is to account for network autocorrelation and spurious correlations between variables (i.e., confounding effects due to the presence of another variable).

As is evident, Model 4 above (with cosine similarity based on security risk factors) is our best model out of the four. The independent variables representing similarity of firms based on internal control reporting by an independent auditor and senior management were insignificant in all models. One explanation for the insignificance of these variables could be that, as firms' internal control reports show, only five independent auditors exist for the 33 firms in our data. Also, there are no material weakness or significant deficiencies as reported in these reports. I found that in cases of no material weakness or significant deficiency, the specific auditing firm issuing the internal control report does not vary significantly. In fact, they look exactly similar and thus firms which are audited by the same auditor around the same time will have very similar internal control report although they are not breached together. Figure 3.4 shows an internal control auditing report by Ernst & Young LLP for Apple for the year 2009 and AMD for the year 2012. The content of the internal control report from auditor matching exactly is highlighted.

AMD 2012:

We have audited the accompanying consolidated balance sheets of Advanced Micro Devices, Inc. as of December 29, 2012 and December 31, 2011, and the related consolidated statements of operations, comprehensive income (loss), stockholders' equity, and cash flows for each of the three years in the period ended December 29, 2012. Our audits also included the financial statement schedule listed in the Index at Item 15(1). These financial statements and schedule are the responsibility of the Company's management. Our responsibility is to express an opinion on these financial statements and schedule based on our audits. We conducted our audits in accordance with the standards of the Public Company Accounting Oversight Board (United States). Those standards require that we plan and perform the audit to obtain reasonable assurance about whether the financial statements are free of material misstatement. An audit includes examining, on a test basis, evidence supporting the amounts and disclosures in the financial statements. An audit also includes assessing the accounting principles used and significant estimates made by management, as well as evaluating the overall financial statement presentation.

Internal control report from independent auditor (Ernst & Young LLP) for AMD for year 2012

Apple 2009:

We have audited the accompanying consolidated balance sheet of Apple Inc. as of September 26, 2009, and the related consolidated statements of operations, shareholders' equity, and cash flows for the year then ended. These financial statements are the responsibility of the Company's management. Our responsibility is to express an opinion on these financial statements based on our audit. We conducted our audit in accordance with the standards of the Public Company Accounting Oversight Board (United States). Those standards require that we plan and perform the audit to obtain reasonable assurance about whether the financial statements are free of material misstatement. An audit includes examining, on a test basis, evidence supporting the amounts and disclosures in the financial statements. An audit also includes assessing the accounting principles used and significant estimates made by management, as well as evaluating the overall financial statement presentation.

Internal control report from independent auditor (Ernst & Young LLP) for Apple for year 2009

Figure 3.4. Example of a highly overlapped internal control report issued by the same auditor in different years for different firms

I found similar observations for others, although the degree of overlap varies across firms. I further verified this observation by running QAP analysis on these two variables (firms with the same auditor and cosine similarity based on auditor report) and found the correlation between the two variables to be 0.184 and highly significant p-value (p=0.00260).

The QAP correlation between management's report on internal control and independent auditing firm is insignificant (p=0.38072) with very low correlation (-0.02). This result is possible because each firm's senior management will approve this report independently from the auditor. That approval is unnecessary in the case for the other internal control variable, which was completely based on auditor report (and was thus highly similar across firms).

In conclusion, the four main independent variables of firm similarity initially proposed in this study that I hypothesized would have positive correlation with breaching together are as follows: 1) business description; 2) security risk factors; 3) similarity based on internal control reporting; and 4) R&D intensity. Of these four, I found support for three for different variations of our model.

3.7 Discussion

This study employs a unique data science approach to exploring information security breaches at the organizational level. The dyadic and the social network viewpoints are novel, as no prior studies have used such techniques to analyze security breaches. This study is an important first step towards understanding the reasons for concurrent breaches. I explore the question of whether firms that are breached together share common attributes. This significant research question needs urgent attention from cyber-security researchers.

Methodologically, this study demonstrates the relevance of specific text mining techniques, including topic modeling and cosine similarity of firms based on their business descriptions, as they can be applied to cybersecurity breaches. The increasing availability of textual data in both structured and unstructured forms provides a unique opportunity to information systems researchers to use text analysis techniques to answer research questions in diverse contexts. Additionally, this study exemplifies the usefulness of data analytic methods as they can be applied to the context of correlated risks in IS. My collection of firm-level characteristics as antecedents that contribute to quantification of correlated risks paves the way for future research of this kind.

Furthermore, I am hopeful that my application of a social network analysis technique, Quadratic Assignment Procedure, for statistical analysis of network-based data will encourage other information systems researchers to use this under-utilized technique. For instance, analysis of online social networks with large amounts of user-generated texts provide a suitable avenue for applying both text analysis and social network analysis techniques such as topic modeling and QAP.

Cyber-security involves securing organizations against security breaches. It is evident from numerous surveys and studies that although businesses have been increasingly spending millions of dollars with each passing year on securing their information infrastructure, many of them invariably experience security breaches. From a practical viewpoint, firms can always learn from the failure of other firms. For instance, a business can take proactive measures in securing their resources if a similar firm has been breached recently. This study demonstrates that business proximity between firms increases the likelihood of breached should either firm experience compromise. Chief Security Officers (CSOs), managers, and employees responsible for safeguarding organizational resources should not only upgrade their software to protect against potential vulnerabilities, but they must also remain informed about breaches affecting other businesses.

From the perspective of cyber-insurance, our approach may be of practical importance to cyber risk actuaries and insurance firms responsible for backing the information infrastructure of multiple firms. Similar firms may have similar portfolios or characteristics, so multiple firms could be at risk of breach. Our approach may assist cyber risk insurers to not only effectively design insurance products for businesses based on their similarity but may also help in computing premiums based on breaches that have affected similar organizations.

The social networking perspective employed in this study is useful in understanding interdependent security vulnerabilities. Although centrality measures did not play a role in our study, they can be insightful in the context of interdependent firms (or interdependent systems within the same firm). For example, a firm with high degree centrality (in terms of its system network connections with others) should have proper security policies in place to ensure that firms they are connected to through their systems are not affected when they are breached. Understanding the network characteristics will enable CSOs and security managers to anticipate the impact that any given node (i.e., firm/business) will have on the entire network. Furthermore, such an understanding will help security experts decide where to deploy resources to mitigate the risks of cyber-attacks or to stem the spread of, for instance, a virus through the network.

As pointed out previously, this study is a preliminary investigation of security breaches from the perspective of business relatedness, which is an antecedent to breach proximity and should be explored using different variables. Firms' utilizations of similar security technologies is an important antecedent to their breach proximity because that similarity increases proximate vulnerability. Lack of availability of such data prevented me from using such variables in my analysis, which comprises only 33 public firms. Although some of the previous MR-QAP studies use similar or smaller network sizes (Coelho et al., 2015; Tsai & Ghoshal, 1998), my professional research plans include replication of this study using a larger sample size. Furthermore, I anticipate extending my research to private firms as well.

3.8 Conclusion

In conclusion, this chapter provides an essential first step towards analyzing security breaches at the dyadic level for organizations that share certain characteristics. Though exploratory in nature, this study proves that the techniques of data science, such as topic modeling, cosine similarity on textual contents, and statistical techniques like QAP for networked data, can be successfully applied in the context of security breaches. I found that for a sample of networked data comprising 33 public firms, the business similarity of firms based on their 10-K filings is significantly correlated with their propensity to be breached together. Also, the similarity of firms based on their security risk factors, R&D spending, and 2-digit SIC code are significantly correlated with the likelihood of being breached together. This study can help researchers and practitioners better understand information security breaches from this unique perspective of business relatedness or similarity.

Appendix B

	SIC code	Headquartered State	Headquartered US Region
ATandT	4812	Texas	South
AdvancedMicroDevices(AMD)	3674	California	West
AppleInc.	3663	California	West
ArcWorldwide	3490	California	West
AutomaticDataProcessing	7374	New Jersey	Northeast
BankofAmerica	6020	North Carolina	Northeast
CapitalOne	6141	Virginia	South
Chevron	2911	California	West
CitigroupInc.	6199	New York	Northeast
DunandBradstreet	7323	New Jersey	Northeast
Facebook	7370	California	West
Google	7370	California	West
HealthNetInc.	6324	California	West
HewlettPackard	3571	California	West
IBM2010	7370	New York	Northeast
IBM2011	7370	New York	Northeast
J.P.MorganChase	6020	New York	Northeast
KrollBackgroundAmerica	6411	Tennessee	South
LinkedIn	7371	California	West
McDonald's	5812	Illinois	Midwest
McKesson	5122	California	West
Nvidia	3674	California	West
RRDonnelley	2750	Illinois	Midwest
Shell	2911	Texas	South
SonyElectronics	3600	New York	Northeast
Symantec	7372	California	West
TeleTech	7389	Colorado	West
ThePrudentialInsuranceCompanyofAmerica	6311	New Jersey	Northeast
Twitter	7370	California	West
USAirways	4512	Texas	South
Unisys	7373	Pennsylvania	Northeast
UnitedHealthcare	6324	Connecticut	Northeast
Yahoo	7370	California	West

Table B.1 Firms breached together with SIC codes, headquartered state and headquartered US region

Firm 1	Firm 2	Firm 3	Firm 4	Firm 5	Firm 6	Date the breach made public
Hewlett Packard	Symantec					11-Dec-08
TeleTech	Sony Electronics					21-Jun-10
Shell *	Chevron					9-Mar-11
McDonald's	Arc Worldwide	IBM2010				14-Dec-10
Citi group Inc.	Bank of America					18-Aug-11
HealthNet Inc.	IBM2011					15-Mar-11
J.P. Morgan Chase	Capital One					12-Feb-13
The Prudential Insurance Company of America	Unisys					4-Mar-13
Advanced Micro Devices (AMD)	Nvidia					13-Jan-13
Apple Inc.	AT and T					9-Jun-10
US Airways[¥]	McKesson	Automatic Data Processing				30-Jul-13
Automatic Data Processing	Facebook	Google	LinkedIn	Twitter	Yahoo	4-Dec-13
Dun and Bradstreet	Kroll Background America[§]					26-Sep-13
RR Donnelley	UnitedHealthcare					28-Jan-13

Table B.2. Adjacency list for the breached network

* Shell is as Royal Dutch Shell with 20-F filing, not 10-K filing

[¥] US Airways was merged with American Airlines in December 2013

[§] Parent organization Marsh & McLennan is used

	Un-Stdized	Stdized Coef	P-value	As Large	As Small	As Extreme	Perm Avg	Std Err
	------------	------------	------------	------------	------------	-------	------------	------------
33 nodes cosine matrix on preprocessed firm descriptions from 10K	0.02488	0.06473	0.02879	0.02879	0.97141	0.04499	0.00016	0.01272
33 nodes risk_factors_cosine_matrix_from_10K	0.0326	0.03531	0.10558	0.10558	0.89462	0.23355	0.00003	0.02873
33 nodes report of independent auditor on IC matrix	-0.01281	-0.02021	0.19716	0.80304	0.19716	0.43651	0.00056	0.01832
33 nodes number of employees from WRDS categorical matrix median cutoff	-0.01116	-0.0409	0.10398	0.89622	0.10398	0.20276	-0.00013	0.00877
33 nodes SIC matrix 2 digit	0.01962	0.05108	0.06139	0.06139	0.93881	0.09938	-0.00009	0.01237
33 nodes R&D to sales matrix previous year median cutoff	0.01317	0.04828	0.06479	0.06479	0.93541	0.13237	0.00004	0.00882
33 nodes HQ_region matrix	-0.0117	-0.04018	0.10438	0.89582	0.10438	0.21516	0.00013	0.00956
33 nodes management_report_on_IC matrix	-0.00196	-0.00553	0.41892	0.58128	0.41892	0.86043	0.00017	0.01154
Intercept	-0.00542	0	0	0	0	0	0	0

MR-QAP analysis result Model (1)

	Un-Stdized	Stdized Coef	P-value	As Large	As Small	Extreme	Perm Avg	Std Err
	------------	------------	------------	------------	------------	-------	------------	------------
33 nodes R&D to sales matrix previous year median cutoff	0.01335	0.04895	0.06399	0.06399	0.93621	0.12977	0.00004	0.00892
33 nodes SIC matrix 2 digit	0.01827	0.04755	0.07339	0.07339	0.92681	0.12278	-0.00008	0.01218
33 nodes report of independent auditor on IC matrix	-0.01197	-0.01889	0.22136	0.77884	0.22136	0.47111	0.00014	0.01848
33 nodes risk_factors_cosine_matrix_from_10K	0.03154	0.03416	0.12318	0.12318	0.87702	0.25255	0.00006	0.02871
33 nodes cosine matrix on preprocessed firm descriptions from 10K	0.02529	0.0658	0.0224	0.0224	0.9778	0.03799	-0.00011	0.01263
33 nodes HQ_state matrix	-0.01057	-0.03129	0.15637	0.84383	0.15637	0.31954	0.00015	0.01071
33 nodes management_report_on_IC matrix	-0.00448	-0.01263	0.33433	0.66587	0.33433	0.68166	-0.00008	0.01107
33 nodes numberr of employees from WRDS categorical matrix median cutoff	-0.01155	-0.04234	0.08758	0.91262	0.08758	0.17676	0.00018	0.00866
Intercept	-0.00526	0	0	0	0	0	0	0

MR-QAP analysis result Model (2)

	Un-Stdized	Stdized Coef	P-value	As Large	As Small	Extreme	Perm Avg	Std Err
	------------	------------	------------	------------	------------	-------	------------	------------
33 nodes R&D to sales matrix previous year median cutoff	0.01344	0.04929	0.06059	0.06059	0.93961	0.12557	-0.00009	0.00876
33 nodes SIC matrix 2 digit	0.01881	0.04897	0.06919	0.06919	0.93101	0.11238	0.0002	0.01227
33 nodes report of independent auditor on IC matrix	-0.01162	-0.01834	0.21936	0.78084	0.21936	0.46931	0.00003	0.01814
33 nodes risk_factors_cosine_matrix_from_10K	0.02954	0.03199	0.13917	0.13917	0.86103	0.29094	-0.00041	0.02911
33 nodes cosine matrix on preprocessed firm descriptions from 10K	0.0249	0.06478	0.02879	0.02879	0.97141	0.04919	-0.00015	0.01284
33 nodes HQ_state matrix	-0.01122	-0.0332	0.13577	0.86443	0.13577	0.26855	-0.00008	0.01036
33 nodes number of employees from WRDS categorical matrix median cutoff	-0.01172	-0.04298	0.07938	0.92082	0.07938	0.16977	0.00012	0.00861
Intercept	-0.00629	0	0	0	0	0	0	0

MR-QAP analysis result Model (3)

	Un-Stdized	Stdized Coef	P-value	As Large	As Small	Extreme	Perm Avg	Std Err
	------------	------------	------------	------------	------------	-------	------------	------------
33 nodes security_risk_factors cosine matrix	0.03583	0.05358	0.03319	0.03319	0.96701	0.08078	0.00004	0.02076
33 nodes R&D to sales matrix previous year median cutoff	0.01267	0.04644	0.07359	0.07359	0.92661	0.14117	0.00011	0.00867
33 nodes number of employees from WRDS categorical matrix median cutoff	-0.01125	-0.04125	0.09638	0.90382	0.09638	0.19296	-0.00001	0.00858
33 nodes SIC matrix 2 digit	0.02087	0.05432	0.05159	0.05159	0.94861	0.07818	0.00014	0.01213
33 nodes report of independent auditor on IC matrix	-0.01349	-0.02128	0.19256	0.80764	0.19256	0.41232	0.00015	0.01847
33 nodes management_report_on_IC matrix	-0.0046	-0.01297	0.34353	0.65667	0.34353	0.67826	-0.00018	0.01157
33 nodes cosine matrix on preprocessed firm descriptions from 10K	0.02492	0.06485	0.0214	0.0214	0.9788	0.03739	-0.0001	0.01241
33 nodes HQ_region matrix	-0.01182	-0.04057	0.10818	0.89202	0.10818	0.21316	-0.00024	0.00967
Intercept	-0.00255	0	0	0	0	0	0	0

MR-QAP analysis result Model (4)

Figure B.3 MR-QAP analysis results for model (1), model (2), model (3) and model (4)

Chapter 4 Antecedents of Insider Security Breaches: An Employee's Perspective

4.1 Abstract

Insider security breaches carried out by current and/or former employees of organizations are of growing concern to businesses. This chapter provides empirical evidence to support both a needs satisfaction perspective and a motivation theories perspective towards understanding insider cybersecurity breaches. To make data-driven conclusions about employee inclinations to commit cybersecurity breaches, the research in this chapter employs a tone analysis service provided by IBM Watson's Tone Analyzer[36] alongside logistic regression analysis. Data for the study were obtained from secondary sources, including Glassdoor,[37] a well-known job search and employer review website. This chapter's objective is to investigate the nature of employees' perceptions towards employers when insider breaches occur. The theoretical paradigms at the foundations of this chapter include the conservation of resources theory, social bonding theory, and theories of motivation from organizational psychology. Matched sample comparison group methodology on a sample of 71 public firms (41 breached and 30 non-breached) was used to carry out the analysis. The results of logistic regression show that the overall rating and rating for compensation and benefits for firms are significantly correlated with insider breaches. An analysis of the text of employees' reviews using IBM's Tone Analyzer and Linguistic Inquiry and Word Count (LIWC) software revealed that emotional tones of joy, fear, and anxiety are significant predictors of an insider breach. The study has strong implications for both theory and practice.

[36] https://www.ibm.com/watson/services/tone-analyzer/
[37] https://www.glassdoor.com/Reviews/index.htm

4.2 Introduction

The final essay in this dissertation studies a specific category of security breaches known as insider breaches. Insider security breaches are a class of breaches carried out by present or past employees of an organization. In this study, insider breaches are regarded as those that are intentionally committed by employees. Because unintentional insider breaches, like those possible when accidents make firms vulnerable to hacking or those that occur because of employees' compromised computers, lack a conscious objective, I do not consider them in this study. According to the Computer Emergency Response Team or CERT insider threat center at Carnegie Mellon University's Software Engineering Institute, an insider threat is defined as "a current or former employee, contractor, or other business partner who has or had authorized access to an organization's network, system, or data and intentionally misused that access to negatively affect the confidentiality, integrity, or availability of the organization's information or information systems."[38] Such threats from insiders include espionage, fraud, information theft, and the theft of intellectual property.

According to the US state of cybercrime survey of 2016, 27% of electronic crimes either were caused or were suspected to have been caused by insiders.[39] Figure 4.1 shows the number of insider breaches in the US over the past 10 years across different industries based on the Privacy Rights Clearinghouse breach dataset used in this study. The abbreviations used in the figure are as follows: BSF indicates businesses offering financial and insurance services, BSO refers to other businesses; BSR indicates retail (both online and offline) businesses; EDU is used for educational businesses; GOV for government; MED for healthcare and medical insurance; and NGO denotes non-profit institutions.

[38] https://insights.sei.cmu.edu/insider-threat/2017/03/cert-definition-of-insider-threat---updated.html
[39] https://insights.sei.cmu.edu/insider-threat/2017/01/2016-us-state-of-cybercrime-highlights.html

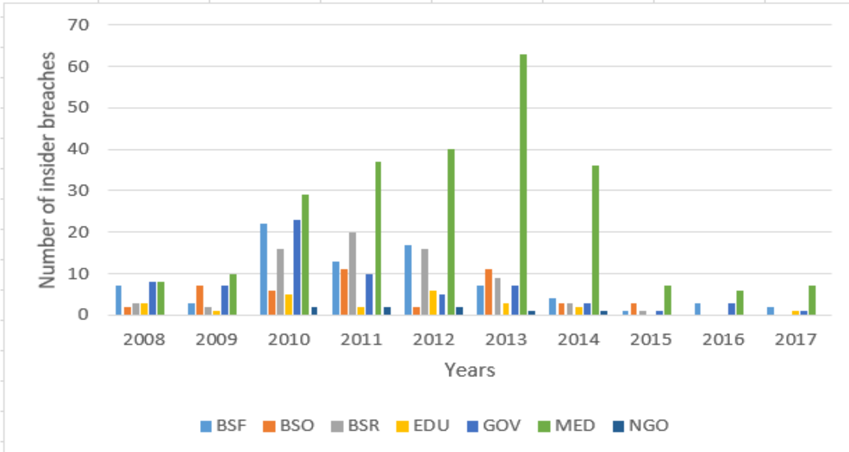


Figure 4.1. Insider breaches over the last 10 years

Several IS studies contend that spending on security technology alone may not help safeguard against breaches when employees are not compliant with organizational security policies (Vance, Siponen, & Pahnila, 2012; Vroom & von Solms, 2004). Similarly, some researchers conclude that insiders comprise the greatest sources of vulnerability and are in fact the weakest links in the compromising of organizational security (Bulgurcu, Cavusoglu, & Benbasat, 2010).

This study investigates insider security breaches from the perspective of employees' motivations and senses of satisfaction or dissatisfaction with their employers. These employee perspectives or sentiments provide the bases for some of the factors underlying insider breaches. The theoretical backdrop for this study includes motivation, conservation of resources, and social bonding (for workplace deviant behavior). Data were gathered from two publicly available sources: 1) the insider breach data for the last decade (2008–2017) from the data breach dataset compiled by a non-profit organization, Privacy Rights Clearinghouse[40]; and 2) data regarding employees' perceptions of the employer from Glassdoor, a well-known job-search and employee review website.[41] I analyze the content of Glassdoor company reviews and ratings for compensation and benefits, and I also interpret and analyze various emotional tones embedded in the textual

[40] https://www.privacyrights.org/data-breaches
[41] https://www.glassdoor.com/Reviews/index.htm

content of reviews, which contain both advantages and disadvantages for employees (the latter of which are potential antecedents of insider breaches).

This study is the first to examine empirically the link between insider breaches and employees' perceptions of their employer organizations. Specifically, I investigate whether employees' ratings of their firms and their perceptions as reflected in textual reviews are correlated with the subsequent realizations of security risk in the form of insider breaches. I expect firms with lower employee ratings and negative textual reviews to have higher occurrences of insider breaches. Researchers and practitioners better understand the causes of insider breaches when they pay attention to the perceptions and sentiments of insiders themselves (employees) regarding their organizations/employers from a human behavior perspective. As a starting point or as a feedback mechanism, conclusions from this chapter can help cybersecurity managers design organizational policies and practices that will reduce insider threats.

4.3 Theoretical Background

This section discusses theoretical perspectives that that can be used to explain the occurrence of insider breaches.

4.3.1 Conservation of resources theory

According to the conservation of resources theory, resources are objects, personal characteristics, conditions, or energies valued by an individual (Hobfoll, 1989). Psychological stress arises from environments in which exist either threats of net loss or actual net loss of such resources or lack of resource gain following individuals' investments in said environments (Hobfoll, 1989). Thus, environmental circumstances like work-life conflict may lead to loss of implied resources or may make those resources difficult conserve, so those circumstances lead to stress. Similar to the coping model (Folkman & Lazarus, 1988; Folkman, Lazarus, Dunkel-Schetter, DeLongis, & Gruen, 1986), the conservation of resource model posits that when individuals are confronted with stress, they strive to minimize resource loss. According to (Hobfoll, 1989), to "offset resource loss, individuals often invest their time and energy in attempts to translate them to other highly prized resources, for example, power and money." Since financial frauds such as theft of credit card information, identity theft, and theft of intellectual property are important factors driving insiders towards breach commitments, the conservation of resources theory can successfully be used as the basis to explain the theoretical and practical reasoning behind employees' motivations to commit insider breaches. The conservation of resources theory also proposes the concept of loss spirals, which entail a subsequent loss of resources by individuals who currently lack resources (Hobfoll, 1989). Such individuals often attempt to capitalize on the resources they have, but that process of capitalization often culminates in self-defeat—or, a further decrease in the likelihood of gaining access to other resources. The concept of loss spiral from the conservation of resources theory has been tested empirically (Demerouti, Bakker, & Bulters, 2004; N. M. Heath, Hall, Russ, Canetti, & Hobfoll, 2012).

4.3.2 Social bonding theory

Social bonding theory is a well-known sociological theory that explains delinquent behavior among individuals in society. This theory, which was later developed into the social control theory as a general theory of crime (Gottfredson & Hirschi, 1990), was originally proposed by (Hirschi, 1969). A social bond is the connection between an individual and society (Shoemaker, 2018). According to social bonding theory, deviance or delinquency occurs when social bonds are lacking (Wiatrowski, Griswold, & Roberts, 1981) or absent (Durkin, Wolfe, & Clark, 1999).

Attachment, commitment, involvement, and belief, are the four primary elements of the social bond, all of which elements have a negative relationship with delinquent behavior (Hirschi, 1969). Thus, the stronger these elements for an individual, the less likely it is that she will be involved in criminal or delinquent behavior. Attachment refers to the social bonding between an individual and significant other, such as family members, friends, and community institutions. From the social norms perspective, individuals with strong and stable attachments with other members of society are less likely to exhibit the delinquent behavior that can cause violations of social norms (Hirschi, 1969). An individual may also fear the breach of social relationships with significant others and therefore may refrain from delinquent behaviors. According to (Hirschi, 1969), commitment refers to investments in the forms of time, money, or efforts made by an individual towards some cause. Individuals with significant investments in certain commitments (such as family relationships or college education) are hesitant to violate their own investments and will be less likely to engage in deviant behaviors in comparison with others who made no such investments. The involvement component of social bond concerns the time spent by an individual on daily activities leading to success and attainment of status objective (Wiatrowski et al., 1981). The involvement component assumes that a person who is more involved in professional activities will have less time for deviant behavior. The final element of social bond theory is an individual’s belief in the moral value systems, social rules, and respect for authority (Durkin et al., 1999; Hirschi, 1969). According to (Hirschi, 1969), following moral beliefs and guidelines set forth by society decreases one’s deviance from social norms and from delinquent behaviors. The individual's attachment to social moral codes depends on the cultural or social norms of a society, and on the individual's perception of society and others in his or her communities. Secondary data sources include a business's "culture and values" rating component, which represents the extent to which an employee finds accordance with organizational norms. However, I could not include that component in this chapter because the "culture and values" rating was incorporated into Glassdoor after June 2012, and the majority of firms in our dataset experienced insider breaches before June 2012.

4.3.3 Need satisfaction perspective

(Sirgy, Efraty, Siegel, & Lee, 2001) provide a new measure for quality of work-life balance based on the need-satisfaction perspective and spillover theories, which can be used to explain the behavior of employees committing breaches within organizations (F. I. Herzberg, 1966; Maslow, 1943; McClelland, 1967; Pardee, 1990).

In order to measure the quality of work-life balance from a needs satisfaction perspective, researchers assume that individuals have some basic needs in life which they aim to fulfill through work (Sirgy et al., 2001). Based on the need satisfaction theories, (Sirgy et al., 2001) identify seven major employee needs: 1) health and safety needs; 2) economic and family needs; 3) social needs; 4) esteem needs; 5) actualization needs; 6) knowledge needs; and 7) aesthetic needs. Furthermore, "employees derive satisfaction from their jobs to the extent that their jobs meets these needs" (Sirgy et al., 2001). That is to say, employees have certain basic needs in life, such as economic needs and/or knowledge needs, which they try to fulfill through their jobs. If fulfilled, an individual's work satisfies her. In the context of this study, economic needs and survival needs are the primary motivators for insider breaches, as when employees steal credit card information, consumer identities, and/or intellectual property to sell to competitors. Needs satisfaction theories (also known as theories of human motivation), such as those based on Maslow's hierarchy of needs, Herzberg's theories of the two-factor model with hygiene factors, and McClelland's need for assessment theory can be drawn upon to explain the behavior of employees who commit insider breaches.

4.3.4 Motivation Theories

4.3.4.1 Maslow’s hierarchy of needs

According to the Maslow’s hierarchy of needs (Maslow, 1943), human needs arrange themselves in a hierarchy of predominance with the satisfaction of higher needs in hierarchy resting upon the prior satisfaction of lower level needs. (Maslow, 1943) states, “a want that is satisfied is no longer a want. The organism is dominated by and its behavior organized only by unsatisfied needs.” According to the needs hierarchy, no rigid arrangement of needs exists, so some individuals value a lower level need more than needs higher up in the hierarchy. Also, multiple needs may exist at the same time with relative partial satisfaction or dissatisfaction based on cultural specifications (Maslow, 1943). The motivational determinants for needs may also lead to satisfaction of more than one need rather than a single need.

Maslow’s hierarchy theorizes five categories of human needs: physiological, safety, love, esteem, and self-actualization. Physiological needs are the basic human needs for survival—food, water, sleep, sex, shelter, sensory pleasures, etc. (Pardee, 1990). As per Maslow, these physiological needs are the most predominant of all needs and the most important human motivations will fulfill physiological needs before any other. Once basic needs are satisfied, higher needs emerge. Safety needs are the next group to emerge. According to Maslow, those with safety needs will try to organize their environments so that anything unexpected cannot occur (Maslow, 1943). Protections against danger and threat are examples of safety needs. In an organizational setting, employment uncertainty, unpredictable policies, favoritism, or nepotism are motivators of safety needs (Hamner & Organ, 1978). Love needs include feelings and expressions of affection and belonging, and they are are the subsequent emergent group of needs. Esteem needs entail desires for self-esteem and achievement recognition. The desire for achievement, freedom, confidence, reputation, prestige, importance, or appreciation are some examples of esteem needs of humans (Maslow, 1943). Lack of satisfaction of these needs may lead to feelings of depression or worthlessness. Finally, self-actualization is the individual’s awareness of the self’s potential for achievement of self-fulfillment

(Maslow, 1943). The self-actualization need later became the basis for another popular theory of motivation known as theory X and theory Y by (McGregor, 1960).

Although not specifically mentioned in the original theory of motivation, recompense for work is the outcome of employment that falls under the safety needs of individuals. If that need for safety is unmet in a work environment, monetary gain tends to arise as one of the primary motives for employees who commit insider breaches. (Sirgy et al., 2001) include pay as one of the safety needs in their development of the measure of the quality of work-life balance. From an organizational psychology perspective, as long as employment includes elements satisfying higher-order human needs, then employment has the potential to motivate employees (Pardee, 1990). The motivational process for fulfilling needs is akin to a decision-making process, and the individual's attitude within her surrounding environment will determine the route she takes to fulfill those needs (Aldag & Brief, 1979; Haimann, 1973).

2.3.4.2 Herzberg's two-factor theory of motivation and hygiene factors

Herzberg's theory of motivation designates two factors—motivation and hygiene—that affect job satisfaction or dissatisfaction, respectively (F. I. Herzberg, 1966; F. Herzberg, Mausner, & Snyderman, 1978; Pardee, 1990). Examples of motivating factors affecting job satisfaction are achievement, recognition, growth, advancement, and responsibility; alternatively, hygiene factors responsible for job dissatisfaction are organizational policy, supervision, work conditions, salary, personal life, job security, and collegiality (F. Herzberg et al., 1978; Pardee, 1990). Literally, hygiene indicates practices conducive to maintaining health and preventing diseases, especially through cleanliness. In turn, Herzberg reformulated the phrase to represent factors responsible for removal of hazards or avoidance of pain in an organizational environment (Duttweiler, 1986; Pardee, 1990). In the two-factor theory, motivational and hygiene factors differ in fundamental ways (F. Herzberg, 1976; Pardee, 1990). For instance, improving hygiene in an organizational setting has short-term effects while increasing motivators has long-term effects. Also, motivators are additive in nature while hygiene factors are cyclical.

According to this theory, hygiene factors—although as important as motivating factors—do not necessarily motivate employees and may even lead to negative effects over the long run (F. I. Herzberg, 1966; F. Herzberg et al., 1978). This theory implies that job satisfaction and job dissatisfaction are compatible constructs—i.e., absence of job dissatisfaction does not inherently indicate job satisfaction (F. Herzberg, 1976). Thus, true job satisfaction involves improving motivating factors, which improvement requires more organizational efforts than required for the expansion of hygienic factors. In other words, hygiene factors provide a base to avoid job dissatisfaction but do not necessarily lead to higher job performance or job motivation. According to Herzberg, motivating factors leading to job satisfaction are intrinsic, while factors leading to job dissatisfaction are extrinsic hygiene factors. Also, extrinsic factors are more easily malleable than the more complex and intrinsic motivational factors, which lead to job satisfaction. However, Herzberg's two-factor theory states that motivational factors will not lead to job satisfaction if extrinsic factors such as salary or working conditions are weak or absent. Thus, if employee dissatisfaction is a major problem within the organization, then managers should direct efforts towards improving hygiene factors.

As per Herzberg's two-factor theory, employee compensation can be considered both as a motivator and a hygiene factor (F. Herzberg, 1976). As previously stated, within the context of this study, perceived imbalances in money or monetary benefits are the most important factors for employees committing insider breaches. (Hersey et al., 2007; Pardee, 1990) provide a comparison of Maslow's need hierarchy with Herzberg's two-factor theory, which are summarized in Figure 4.2:

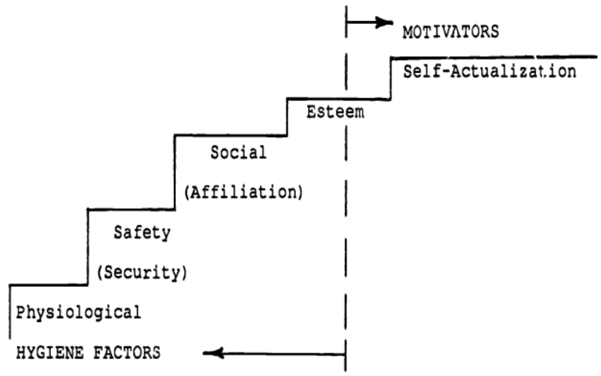


Figure 4.2 Relationship between Herzberg's two-factor theory and Maslow's need hierarchy (Hersey et al., 2007; Pardee, 1990)

According to Hersey et al. (2007), Maslow's need hierarchy helps identify needs while Herzberg's theory provides scholars with insights in the form of motivations and hygiene factors responsible for satisfying those needs. As depicted in Figure 4.2, hygiene factors from the two-factor theory overlap with the lower or extrinsic needs from Maslow's need hierarchy, whereas motivating factors leading to job satisfaction from need hierarchy correspond with the higher-order need of self-actualization from Maslow.

2.3.4.3 McClelland's Need for Achievement Theory

(McClelland, 1967) theorizes the need for achievement—the n-Ach theory, in popular parlance—and inasmuch provides researchers with a similar need-based and motivational perspective towards explaining the motivations of employees committing insider breaches. According to McClelland's theory, the strong needs of an individual motivate her to behave in ways associated with satisfying those needs (Pardee, 1990). One of the factors reflecting the need for achievement involves achievers taking calculated risks (Pardee, 1990), which are akin to committing insider breaches in hopes of avoiding detection and retribution. This theory presupposes that within a particular environment, needs are learned through coping mechanisms (Pardee, 1990). In the case of an insider breach, this theory implies that organizational policies, such as that governing non-compliance, will influence employees committing breaches. Although this study does not investigate the question of whether a lack of strong security compliance policy leads to more insider breaches, the reality of this positive relationship is worth investigating. For instance, (Greenhaus & Beutell,

1985) propose that “work-family conflict is strongest when there are negative sanctions for noncompliance with role demands.” In other words, a lack of strong sanctions or negative sanctions for noncompliance for a specific job role may be positively associated with work-life conflict. Although (Greenhaus & Beutell, 1985) do not provide empirical support for this claim, research on this question may provide useful insight into insider breaches.

4.3.5 Hypothesis Development

4.3.5.1 Compensation & benefits and overall Glassdoor rating

Past research on organizational commitment has revealed a positive relationship between job satisfaction, organizational commitment, and employees’ adherence to organizational policies (Ingersoll, Olsan, Drew-Cates, DeVinney, & Davies, 2002; Tett & Meyer, 1993; Williams & Anderson, 1991). Lack of career advancement opportunities and low salary benefits were found to be responsible for higher job dissatisfaction (Igbaria & Guimaraes, 1993). One popular construct that measures an employee’s feelings regarding her job was designed and developed as the Job Descriptive Index (JDI) by (P. C. Smith & others, 1969). The JDI was designed along five major sub-dimensions or components of job satisfaction: satisfaction with work, supervision, coworkers or colleagues, pay scale, and promotion. In this chapter, the secondary data source, Glassdoor, enables a measure of pay scale in terms of compensation and benefits. Additionally, job satisfaction is found to have a negative relationship with absenteeism, ill health, and grievance. In the context of this study, lack of job satisfaction may lead to grievance towards an organization (Locke, 1976). Within the context of IS research, relationships between role stressors (role ambiguity and role conflict), job satisfaction, organizational commitment, and intention to quit are investigated by (Baroudi, 1985; Igbaria & Guimaraes, 1993). Both studies found role conflict and role ambiguity to be negatively related with job satisfaction, which in turn is a significant predictor of an employee’s organizational commitment and intention to remain in employment.

Based on the literature, I argue that employees dissatisfied with their work or job roles may be emotionally, mentally, or physically withdrawn from their employers. The literature on organizational commitment and job satisfaction, along with the social bonding theory, suggests that lack of job satisfaction (due to dissatisfaction with compensation and benefits) will lead to less organizational commitment and hence deviant behavior. Thus, in many such cases, employees will quit immediately or after some time; or, they will continue half-heartedly for a while; or, they may commit grievance activities (such as insider breaches) and/or may disobey organizational policies.

Hypothesis 1: Firms with employees with lower satisfaction regarding compensation and benefits are more likely to be victims of insider breaches than firms with employees with higher satisfaction regarding compensation and benefits.

The job website, Glassdoor, is the main source of secondary data in this study. Along with the above ratings compensation and benefits, Glassdoor also provides an overall employer rating for each employee review. Thus, we also hypothesize the following relationship (Hypothesis 2, below) between the overall employer (or firm) rating and its likelihood to be affected by insider breach.

Hypothesis 2. Firms with lower overall ratings from current or past employees are more likely to be victims of insider breaches than firms with higher overall ratings.

4.3.5.2 Textual tones in employee reviews and insider breaches

Analyzing text through techniques of sentiment analysis and opinion mining provides useful insights on the implications of textual content to answer the research question in different contexts. According to (Liu, 2012), sentiment analysis or opinion mining concerns analyzing “people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, topics, and their attributes.” With the proliferation of opinionated content, sentiment analysis and opinion mining have found a wide variety of applications. Some examples include sentiment analysis of Twitter (Pak & Paroubek, 2010), Amazon (Jo & Oh, 2011), real-time sentiment analysis of

Twitter for US presidential elections (H. Wang, Can, Kazemzadeh, Bar, & Narayanan, 2012), stock market prediction based on Twitter sentiments (Bollen, Mao, & Zeng, 2011), detection of opinion spam in online reviews from Amazon.com (Jindal & Liu, 2008), sentiment analysis of blogs and online news articles (Godbole, Srinivasaiah, & Skiena, 2007), and sentiment analysis of movie reviews (Tun Thura Thet, Na, & Khoo, 2010). Other interesting applications of these techniques include sentiment analysis of clinical data in medical settings (Denecke & Deng, 2015), sentiment analysis of suicide notes (Pestian et al., 2012), and sentiment analysis for competitive intelligence (Xu, Liao, Li, & Song, 2011), among others.

The emergence of job websites such as Glassdoor,[42] Great place to work,[43] Indeed,[44] Vault,[45] and Job Crowd[46] allows current and previous employees to write anonymous reviews of employers. These anonymous reviews present unique opportunities to information system researchers to extract employees' perceptions towards employers. Two recent studies have used opiniated textual reviews from Glassdoor to extract employee satisfaction figures and determine those figures' relationships with organizational performance (DeKay, 2013; Luo, Zhou, & Shon, 2016). In this study, I propose that emotional tones, as expressed in employee reviews, of anger, disgust, fear, sadness, and joy are likely to be correlated with insider breaches, but the directions of relationship will differ. I propose that emotional tones of anger, disgust, and sadness have positive relationships with potential for insider breaches, while fear and joy are negatively correlated with breach probability.

Similarly, I propose that analytical and confident language tones are negatively correlated with the likelihood of firms experiencing insider breach. Conversely, tentative language tones should be positively correlated with breach probability.

[42] https://www.glassdoor.com/index.htm
[43] http://reviews.greatplacetowork.com/
[44] https://www.indeed.com/Best-Places-to-Work
[45] http://www.vault.com/rankings-reviews/explore-companies
[46] https://www.thejobcrowd.com/

This research uses a tone analysis service provided by IBM Watson's tone analysis API,[47] callable through Python and other languages, to extract five emotional tones: anger, disgust, fear, joy, and sadness. The emotional tone of anxiety is extracted using the Language Inquiry and Word Count Software (LIWC) (Pennebaker & Graybeal, 2001; Tausczik & Pennebaker, 2010) used in a previous IS study (Yin, Bond, & Zhang, 2013). Psycholinguistics provide the foundation for tone analysis (Fodor, Bever, Garrett, & others, 1974; Massaro, 2014), according to which the usage of certain words by individuals in their daily communications can provide researchers with clues about users' personalities, emotional states, and thought processes. For instance, some studies in psycholinguistics derive five primary personality traits from individuals' social media posts (Gou, Zhou, & Yang, 2014; J. Zhao, Gou, Wang, & Zhou, 2014).

No previous study has explored empirically the possible relationship between the emotional tones expressed by employees while expressing their feelings towards employers and insider security breaches. The purpose of including emotional tones extracted from employee reviews in this research is to explore that possible relationship. Although I have not formally hypothesized the relationships between emotional tones and insider breaches, relevant literature from theories of emotions (Moors, Ellsworth, Scherer, & Frijda, 2013; C. A. Smith & Ellsworth, 1985) suggests the following relationship to be held true: the emotional tone of anger and sadness will have positive relationship with likelihood of firm to be breached, whereas joy, fear, and anxiety will be negatively correlated with the potential of firm to be breached.

This exploration of the relationships between information security breaches and the emotional tones as expressed in reviews of the advantages and disadvantages of employers by current and former employees is the most salient contribution of this study. The application of IBM Watson's API service to extract emotional tones of sadness, joy, fear, and anger from the employee reviews is explained in the section on data collection. As mentioned previously, the emotional tone of anxiety is extracted using Language Inquiry and Word Count software.

[47] https://www.ibm.com/watson/developercloud/tone-analyzer/api/v3/python.html?python

4.4 Data Collection

Data collected for this final dissertation essay is derived from two publicly available sources: data about public companies which were affected by insider breaches in the past 10 years (2008–2017) was collected from the Privacy Rights Clearinghouse's data breach dataset,[48] and data for employees' perceptions towards these firms was gathered from the popular job website Glassdoor.[49] Although the breach dataset contains data about firms that have been breached from 2005 onwards, the job and employee review website, Glassdoor, was founded in June 2007. Therefore, ratings and reviews data commenced in 2008. Hence, the dataset for this chapter ranges from 2008–2017.

Privacy Rights Clearinghouse is a California-based privacy advocacy non-profit organization established in 1992. As mentioned previously, its data breach dataset has been used in many information security research articles, both in the fields of information systems as well as computer science (Acquisti et al., 2006; Avery & Ranganathan, 2016; Cachin & Schunter, 2011; Culnan & Williams, 2009; Gatzlaff & McCullough, 2010; Kwon & Johnson, 2015; Posey Garrison & Ncube, 2011; Sen & Borle, 2015).

Few studies have analyzed employee reviews from a popular website like Glassdoor. Of the exceptions (DeKay, 2013; Ji, Rozenbaum, & Welch, 2017; Lee & Kang, 2017; Luo et al., 2016), none focused on cybersecurity breaches, especially insider breaches. Glassdoor is the second largest job site in the US, after Indeed.[50] The Glassdoor website hosts nearly 57 million employee reviews for more than 770,000 companies worldwide.[51] Secondary data from websites such as Glassdoor and others can be useful in exploring and answering many important research questions related to employee-employer relationships.

A typical employee review posted on Glassdoor is shown in Figure 4.3.

[48] https://www.privacyrights.org/data-breaches
[49] https://www.glassdoor.com/Reviews/index.htm
[50] https://www.indeed.com/
[51] https://www.glassdoor.com/press/facts/

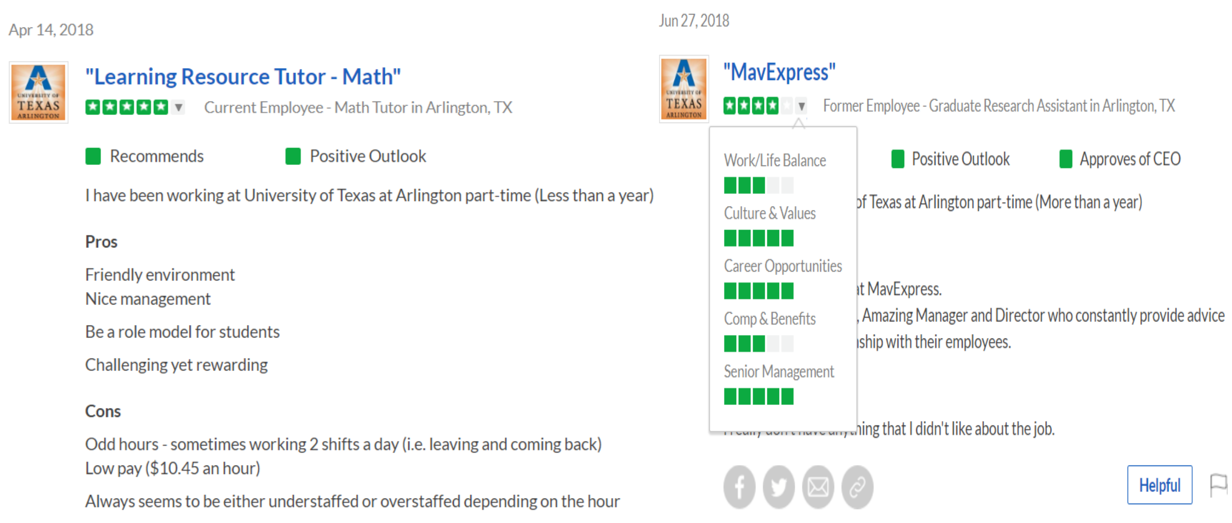


Figure 4.3. Examples of typical employee reviews from Glassdoor

Figure 4.3 shows a typical review on Glassdoor. As is evident, these reviews allow the researcher to extract plentiful information about employees and their perceptions of their employers. Glassdoor reviews include valuable data, including review date; current or former employee status; employee duration with the company (as a dummy, for example, less than a year, more than 10 years, etc.); employee position; approval of employer executives (or not); employee recommendation potential of the employer to a friend; existence of a positive outlook towards the employer (or not); employer advantages; employer disadvantages; advice to management; and ratings for work-life balance, culture and values, career opportunities, and compensation and benefits. The latter five ratings, along with an overall rating, exist on a scale of one (least satisfied) to five (most satisfied). Some employee content fields, such as "advice to management" and "positive outlook," are optional and therefore not considered in this study.

I learned that ratings for "culture and values" were included in Glassdoor around June 2012. Since many of the breaches in my dataset occurred before 2012, I did not include those ratings in this analysis.

This study uses employee review data for 71 firms (41 beached and 30 non-breached) from Glassdoor. The selection of firms and matching of the two groups (breached and non-breached) is explained in the research methodology section.

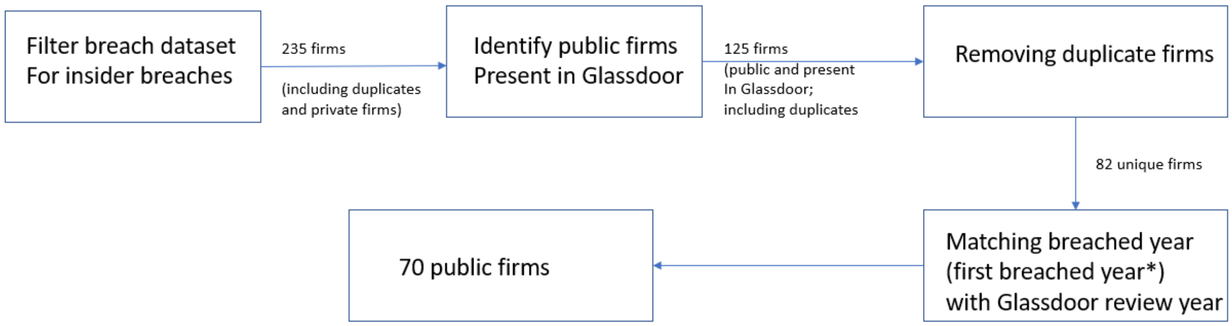
4.5 Research Methodology

Data collection for this study entailed a multi-step process. First, I collected data about firms that have experienced insider breaches and are also listed in the Privacy Rights Clearinghouse's breach data set. This process resulted in 235 firms (including duplicates, as well as private and public firms) which were breached between 2008–2017. Next, I identified public firms and determined those with reviews on Glassdoor. This step resulted in 125 firms which are breached, publicly traded, and present in Glassdoor (but not unique). After removing duplicate firms, I was left with 82 firms.

Since Glassdoor provides review dates, and I wanted to correlate employees' behavioral traits and perceptions towards employers with subsequent insider breach occurrences, my next step in data collection involved matching Glassdoor review dates with the dates when breaches were publicly announced. Thus, I collected Glassdoor reviews and ratings from the first review of the previous year to the date when the breach was announced. For example, if the date for public announcement or disclosure of insider breach was listed as February 23, 2013, then I collected all reviews for that firm starting from the first review in year 2012 up to the last review on or before February 23, 2013. The last review could be February 23, 2013 or sometime in 2012, but no reviews after the reported breach date were collected.

While matching these breached firms to Glassdoor for collecting reviews, I found few firms with no reviews in the breached year. This step left me finally with 70 public firms, which were present in Glassdoor and had at least one review in the breached year. From this list of 70 firms, I further observed a few firms with multiple breach incidents from 2008–2017 (sometimes multiple insider breaches in the same year). In such cases, similar to the approach of (T. Wang et al., 2013), I selected only the first year in which breach occurred. For instance, McDonald's had 10 insider breach incidents in our dataset between the years 2008 and 2012, but I collected reviews only for 2008. Also, since there are no reviews in Glassdoor before June 2008, for firms which were breached in 2008, reviews could only be collected for 2008 until the date when breach was made public.

The multi-process steps involving selection of breached firms is summarized in Figure 4.4.



(a). Selecting firms with insider breaches

Figure 4.4 Selection of firms with insider breaches

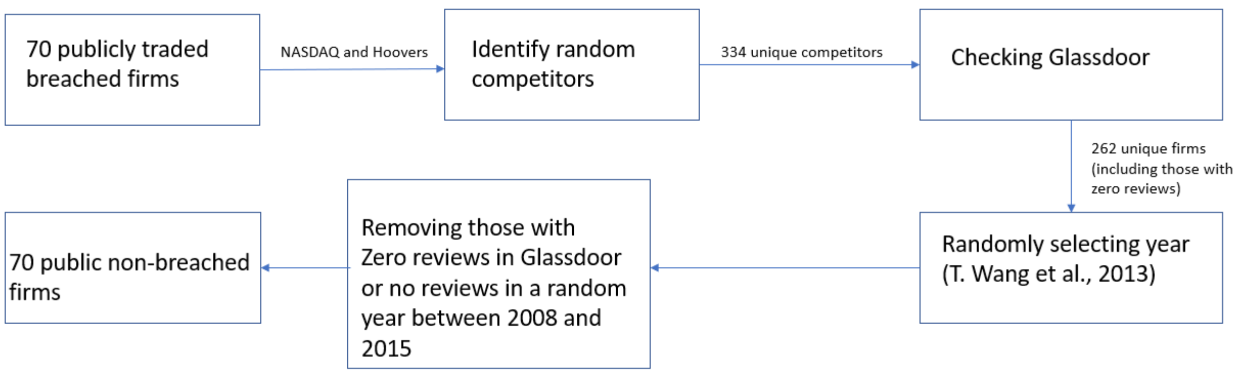
The selection of non-breached firms proceeded in a similar manner, except for the first step wherein I collected a random sample of firms that were competitors of the 70 breached firms. The random sample of competitor firms was collected from either of the two websites, NASDAQ[52] or Hoovers,[53] each of which provides a list of competitor firms based on the input firm. The purpose of this step was to ensure suitable matching between the two groups. I collected 334 unique publicly-traded competitor firms randomly against the 70 breached firms. The next step involved checking how many of these 334 were listed in Glassdoor. This resulted in 262 competitor firms, which are present in Glassdoor, although some have zero reviews.

For selecting the year for non-breached firms, I randomly chose a year between 2008 and 2015. This step of random year selection for non-breached firms is consistent with (T. Wang et al., 2013). For non-breached firms, I collected reviews for approximately one and a half years because I must have approximately the same number of days for reviews collection to ensure balanced comparison between the two groups for the reviews. The average number of days for which reviews were collected for breached firms is 526.96. The

[52] https://www.nasdaq.com/
[53] http://www.hoovers.com/

average number of days for the non-breached 70 firms for which we collected reviews is 486.7. This difference between the average numbers of days for which we collected reviews for the two groups exists because although we collected reviews for one year (365 days), the reviews for the previous year were collected through random dates. Hence, we ended up with a list of 70 non-breached firms, thus giving us equal samples sizes for the breached versus non-breached events, or equal split of 50-50, consistent with the works of (Lancaster & Imbens, 1991; T. Wang et al., 2013).

The multi-process steps involving selection of breached firms is summarized in the Figure 4.5.



(b). Selecting firms with no insider breaches

Figure 4.5. Selection of firms with no insider breaches

The next step in data collection involved gathering data in the forms of ratings and textual reviews from Glassdoor for the two groups, 70 breached and 70 non-breached. During this step, I found that only several firms existed in the two groups with few reviews during the breached and previous year (for breached firms) and a random year (for non-breached firms). The number of reviews for the two groups is available in Appendix C, Tables C.1 and C.2. The joint median for the two groups is 48.5 (40 breached and 31 non-breached), which translates to 56% breached and 44% non-breached firms. Although proceeding with equal ratio for breached and non-breached firms as suggested by (T. Wang et al., 2013) would be ideal, some literature suggests that using unequal ratios of events versus non-events is acceptable for group comparison studies (Breslow, Day, & others, 1980; Pinczowski, Ekbom, Baron, Yuen, & Adami, 1994; Rudolfer, Paliouras, & Peers, 1999).

Next, I collected the firm characteristics of return on assets (ROA), returns on equity (ROE), and earnings per share with the purpose of comparing the two groups (40 breached and 31 non-breached firms) on mean values for these characteristics. Based on the t-test results of non-significant p-values ($p>>0.1$) for three firms' characteristics, I cannot reject the null hypotheses. Therefore, the two groups are not significantly different. In addition, I also performed the Kolmogorov-Smirnov test on the industry distribution of the two groups based on the first two digits of the SIC industry classification codes. This step, comparing 2-digit SIC codes for breached and non-breached firms, is also consistent with (T. Wang et al., 2013). Again, I did not find the two groups to be statistically significantly different; thus, the two samples are derived from the same population. The SIC code industry distribution for the two groups (40 breached and 31 non-breached) is in Table 8 and the complete list of firms (70 breached and 70 non-breached), with their respective SIC codes, number of days for which reviews were collected, and two-digit SIC code description are shown in Tables C.1 (breached) and C.2 (non-breached) in Appendix C.

Breached Firms			Non-breached firms		
Two-digit SIC code	Description	Number of firms	Two-digit SIC code	Description	Number of firms
60	Depository Institutions	10	48	Communications	5
61	Nondepository Credit Institutions	3	73	Business Services	3
48	Communications	3	63	Insurance Carriers	3
73	Business Services	3	61	Nondepository Credit Institutions	2
	Others	21		Others	18

Table 4.1. Industry distribution by 2-digit SIC codes

The process of comparing the two groups (breached and non-breached) is summarized in Figure 4.6.

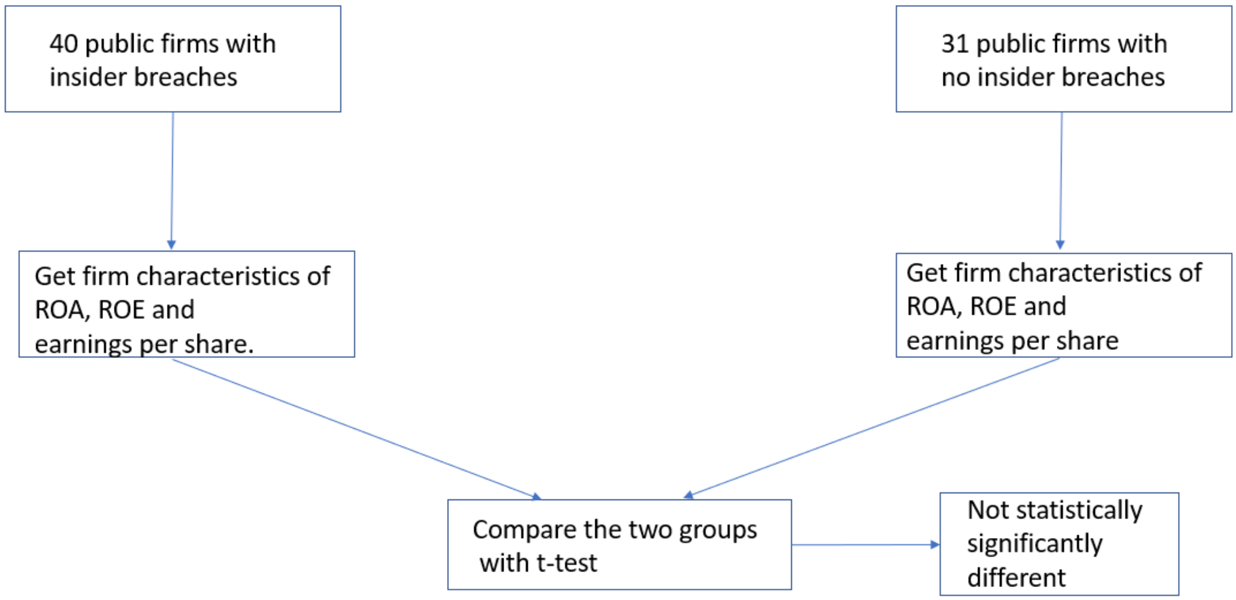


Figure 4.6. Group comparisons (breached vs. non-breached firms) for certain firm characteristics

For each firm, median rating scores were calculated to be used during logistic regression analysis for comparison of the two groups. This practice of choosing median score rather than the average is consistent with many of the previous studies based on online product reviews (Danescu-Niculescu-Mizil, Kossinets, Kleinberg, & Lee, 2009; N. Hu, Liu, & Zhang, 2008; N. Hu, Pavlou, & Zhang, 2006; Ranganathan & Ganapathy, 2002).

As noted, for analysis of textual reviews, I used IBM Watson's Tone Analyzer,[54] which not only provides traditional sentiment scores, such as positive (as joy) and negative (as sadness), but it also assesses other emotions such as anger, fear, and disgust as emotional tones. The science behind IBM's tone analysis service for emotional, language, and social tones is based on emotion analysis from linguistic analysis and psycholinguistics research.[55] When a user signs up for IBM Watson service, she is provided with a username and password, which can then be used to call IBM's tone analysis API from a programming language or

[54] https://www.ibm.com/watson/services/tone-analyzer/
[55] https://console.bluemix.net/docs/services/tone-analyzer/science.html#the-science-behind-the-service

platform. In this study, Python's anaconda platform[56] was used to call IBM's API for tone analysis. The sample Python code for the API call is shown in figure 4.7.

```
import json
from watson_developer_cloud import ToneAnalyzerV3

tone_analyzer = ToneAnalyzerV3(
   version ='2017-09-21',
   username ='{username}',
   password ='{password}'
)

text = 'Team, I know that times are tough! Product sales have been disappointing for the past three quarters. We have a competitive product, but we need to do a better job of selling it!'
content_type = 'application/json'

tone = tone_analyzer.tone({"text": text},content_type)

print(json.dumps(tone, indent=2))
```

Figure 4.7. IBM tone analysis API call example
(Source: https://www.ibm.com/watson/developercloud/tone-analyzer/api/v3/python.html?python#tone)

The output from the above code will be in the form of a JSON file (one JSON per firm), which can be converted to the corresponding CSV file format with all the tonal scores from the textual scores. For further analyses, I used scores for four emotional tones (anger, fear, joy, and sadness) from IBM's Tone Analyzer and a score for anxiety from LIWC. I did not use the emotional tone of confidence as its value is zero for almost all the firms in our dataset. Thus, we have used seven independent variables, based on the tone analysis of employee's textual reviews, to be used during the next step of logistic regression. The descriptive statistics for these independent variables are shown in Table 4.2. Glassdoor ratings are measured between one and five, and tonal scores vary from zero to one (with zero representing least value and one as maximum extracted from review text).

[56] https://anaconda.org/anaconda/python

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
rating_overall_median	71	3.3	0.5	2	3	4	4
rating_comp_median	71	3.38	0.39	3	3	3.5	4
Anger_PC	71	0.13	0.1	0.07	0.09	0.12	0.57
Fear_PC	71	0.1	0.07	0.06	0.08	0.11	0.63
Joy_PC	71	0.61	0.03	0.54	0.59	0.63	0.66
Sadness_PC	71	0.46	0.13	0.16	0.49	0.53	0.57
Anxiety_PC	71	0.25	0.12	0	0.17	0.33	0.7

PC = pros and cons

Table 4.2. Descriptive statistics for Independent Variables

4.6 Analysis and Results

The statistical analysis technique of logistic regression is used when the dependent variable is categorical or binary (as in our case of breached firms—represented as 1—versus non-breached firms—represented as 0), and the independent variables are metric or non-metric. As a statistical technique, the aim of logistic regression is to predict the probability of an event (such as breach) occurring (Hair et al., 2006). The correlation matrices with the corresponding VIF values for the three selected models with highlighted moderate correlations (greater than 0.5) between independent variables are shown in Tables 4.3 and 4.4, respectively.

	rating_overall_median	rating_comp_median	Anger_PC	Fear_PC	Joy_PC	Sadness_PC	Anxiety_PC
rating_overall_median	1						
rating_comp_median	0.394	1					
Anger_PC	-0.093	-0.016	1				
Fear_PC	0.157	0.158	-0.059	1			
Joy_PC	0.296	0.315	0.026	0.098	1		
Sadness_PC	-0.132	-0.157	-0.135	-0.218	-0.28	1	
Anxiety_PC	-0.135	0.005	-0.067	-0.052	-0.096	0.121	1
VIF	1.89257	1.659788	1.461328	2.136791	1.876196	1.066129	3.581502

Table 4.3 Correlation matrix for independent variables used in model (1)

Although the independent variable of anxiety as extracted from the advantages and disadvantages of employee reviews is well below five, it is well above the ideal VIF value of one or below two. Therefore, I removed this independent variable in the third possible model, to see resultant correlations between other independent variables.

	rating_overall_median	rating_comp_median	Anger_PC	Fear_PC	Joy_PC	Sadness_PC
rating_overall_median	1					
rating_comp_median	0.394	1				
Anger_PC	-0.093	-0.016	1			
Fear_PC	0.157	0.158	-0.059	1		
Joy_PC	0.296	0.315	0.026	0.098	1	
Sadness_PC	-0.132	-0.157	-0.135	-0.218	-0.28	1
VIF	1.094491	1.122879	1.004111	1.212834	1.346383	1.24701

Table 4.4 Correlation matrix for independent variables used in model (2)

The output from logistic regression for three useful models is shown in figure 4.7. The complete logistic regression outputs for the two models (with and without the anxiety variable) are shown in Figures C.1and C.2 of Appendix C.

	Dependent variable:	
	Breached_or_not	
	(1)	(2)
Constant	54.901***	32.432***
	(16.055)	(9.799)
rating_overall_median	-2.510**	-1.122*
	(1.146)	(0.626)
rating_comp_median	-3.297**	-1.440*
	(1.380)	(0.829)
Anger_PC	9.220	1.495
	(5.832)	(2.908)
Fear_PC	-106.190***	-33.092*
	(38.916)	(18.997)
Joy_PC	-51.522***	-30.975**
	(18.647)	(12.836)
Sadness_PC	-5.668	-4.217
	(3.521)	(2.565)
Anxiety_PC	29.060***	
	(8.448)	
Observations	71	71
Log Likelihood	-19.895	-36.715
Akaike Inf. Crit.	55.790	87.429
Note:	*p<0.1; **p<0.05; ***p<0.01	
	PC = Pros and Cons	

Figure 4.7. Logistic regression results

Based on the results from the logistic regression, I did find support for majority of independent variables. The effect of overall rating and rating for compensation and benefits on the probability of being breached is statistically significant as well as negative (suggesting that an increase in this rating will decrease the probability of being breached) in all the models. For example, for the model (1), the exponentiated coefficients or odds ratio for the overall median rating is 0.08124006 and 0.3257728 for model (2). Since the original coefficients for independent variable of overall median Glassdoor rating in both model (1) and model (2) are negative, the corresponding values of exponentiated coefficients are less than 1 and the odds of occurrence of breach will decrease with increase in this variable. In terms of percentage change in odds for the overall rating, for model (1) each one unit increase in the ratings will reduce the likelihood of the firm being breached by 91.876% for model (1) and by 67.42% for model (2) (Hair et al., 2006). Similarly, I interpret the results for the effect of other independent variables on the probability of a firm being breached. The significance of compensation and benefits ratings suggest the importance of this variable in reducing insider breaches. The significance of this variable is also in accordance with the theory of conservation of resources, the needs satisfaction perspective, and motivation theories, where pay and employee benefits are a resource or an employee need, which if unfulfilled will lead to an employee committing breach.

Joy or happiness as expressed in the pros and cons part of the textual review is also statistically significant (and negative) in all models. This result suggests that firms with happier employees are less likely to be victims of insider breaches. I also found the variable of fear as extracted from the pros and cons part of the textual review to be significant and negative in model (1). Although the theoretical interpretation of the significance of this variable is difficult to ascertain, in this case, it suggests that firms with more control over employees (resulting in a fearful work environment) are less likely to be breached. This result could be explained from the perspective of deterrent theory and deterrence research from sociology. Usage of deterrent theory in the IS security literature suggests that fear of sanctions (both certain and severe of sanctions) is negatively correlated with deviance behavior (D’Arcy, Hovav, & Galletta, 2009; Detmar W. Straub & Jr., 1990; Silberman, 1976). Results also suggest the significance of anxiety and nervousness, or the negative emotions of tension as expressed in employee reviews. Anxiety as an emotional state arises from unpredictable situations or events rather than from the individual herself (Lerner & Keltner, 2000). Therefore, the positive and significant relationship between anxiety and insider breach warrants further investigation.

4.7 Discussion

This study attempts to provide insight into the motivations that cause insider breaches in organizations. Specifically, the research builds upon the theory of conservation of resources, social bonding theory, workplace deviance, and various motivation theories from organizational psychology to elucidate insider breaches in organizations from an insider's perspective. The main goal of the study was to explore individual level antecedents to insider security breaches, especially those primarily derived from employees' perceptions of their employers. I conclude that an understanding of employees' (or group of employees or team) perceptions of their jobs and their employers will help predict of future breach incidents. This study may help detect and thwart deviant behavior. Although some studies, mainly from government agencies and practitioners, predict insider threats based on employees' perceptions and behaviors, much research remains to be done. For example, from an academic perspective, (Gritzalis, 2014) suggests usage of general deterrence theory (Straub & Welke, 1998), social bonding theory (Gottfredson & Hirschi, 1990; Hirschi, 1969), social learning theory (Bandura, 1978), theory of planned behavior (Ajzen, 1985, 1991), and situational crime prevention theory (R. V. Clarke, 1980; Ronald V. Clarke, 1983), for understanding insider behavior and predicting insider threats or breaches. (Nurse et al., 2014) provides a framework for characterizing cyberattacks with actors' (attackers') personality characteristics and psychological states of the attackers as important components. This study of the first empirical studies that attempts to study insider security breaches from an insider's perspective rather from an organizational perspective.

The relationships hypothesized in this study were tested using secondary data from a unique secondary source, Glassdoor, where employees report how they feel about their jobs and employers. The ratings and reviews, as reported by the current or past employees, are accessible through a third-party job site; hence, neither the organization nor the employer has any influence over these data.[57] That is, the data does not

[57] The companies review section of Glassdoor webpages also states to its users "Your trust is our top concern, so companies can't alter or remove reviews."

suffer from any intrusion by the employer for whom reviews or ratings are posted. Another advantage of using data from a third-party website (instead of survey-based research) is that it does not suffer from response bias, where an employee may report certain ratings or write acceptable reviews having no future consequence.

Reducing insider security breaches is a growing concern for organizations. As per a recent industry report,[58] the average cost of insider security threat in an organization is a staggering $8.7 million. Along with background checks, controlled access, and monitoring user actions, employees' emotions and personality characteristics can be a significant predictor of insider breaches. If employees are unhappy for any number of reasons (low pay, too much work, lack of work-life balance, inept senior management, and so forth), then it is possible that they may end up behaving in a deviant manner or committing a cybersecurity breach. Insider breach is a step-by-step process. For example, a disgruntled employee might perform reconnaissance for identification of data or intellectual property to steal. Hence, early intervention by an organization in terms of mitigating the employee's concerns may reduce the risk of insider breaches. This topic needs further research from both academics and industry professionals.

[58] https://securityintelligence.com/news/the-average-cost-of-an-insider-threat-hits-8-7-million/

4.8 Conclusion

"What motivates employees to commit intentional insider breaches?" is the primary research question addressed in this chapter. Using a sample of breached and non-breached firms, this study relied on several theoretical streams to examine empirically whether employees' perceptions of their firms in addition to their own emotional characteristics affect the likelihood of insider breaches. Based on a unique secondary data source, I found overall Glassdoor ratings and ratings for compensation and benefits by employees to be a significant predictor of insider breaches occurring in public firms. I also found employee happiness/joy, fear, and anxiety when providing pros and cons of their jobs to be significantly correlated with the probability of insider breaches. This research is an initial attempt to view insider breaches from an insider's perspective, and I hope that this work may prove to be a stepping stone towards further empirical research in this domain.

Appendix C

Company	Number of reviews for first breach year	Number of days of reviews	SIC code	two digit SIC code description
AMD	183	379	3674	Electronic & Other Electrical Equipment & Components
American Airlines	61	413	4512	Transportation by Air
American Express	222	592	6141	Nondepository Credit Institutions
Arco Gas Station(BP)	49	533	2911	Petroleum Refining and Related Industries
AT&T	519	555	4812	Communications
Bank of America	438	524	6020	Depository Institutions
Beam Global Spirits & Wine Inc.	3	532	2085	Food and Kindred Products
Bed Bath and Beyond	741	633	5700	Home Furniture, Furnishings and Equipment Stores
Burger King	72	420	5812	Eating and Drinking Places
Capital Grille	5	687	5812	Eating and Drinking Places
Capital One Bank	213	494	6141	Nondepository Credit Institutions
Charter One / Citizens Bank	107	605	6020	Depository Institutions
Cheesecake Factory	27	619	5812	Eating and Drinking Places
Chili's	139	606	5812	Eating and Drinking Places
Citibank	420	652	6199	Nondepository Credit Institutions
Comfort Inn and Suites	46	447	6794	Holding and Other Investment Offices
Court Ventures	108	660	8742	Engineering, Accounting, Research, and Management Services
CVS Caremark Corp.	303	433	5912	Miscellaneous Retail
Duane Reade	19	510	5912	Miscellaneous Retail
Fifth third bank	142	605	6020	Depository Institutions
FirstMerit Bank	3	605	6020	Depository Institutions
Fox Entertainment Group	10	472	4888	Communications
General Communication Inc. (GCI)	8	509	5961	Miscellaneous Retail
General Motors Co.	191	580	3711	Transportation Equipment
H&R Block	28	447	7200	Personal Services
Home Depot	612	713	5211	Building Materials, Hardware, Garden Supplies & Mobile Homes
HSBC Bank USA National Association	812	668	6020	Depository Institutions
Huntington National Bank	20	494	6020	Depository Institutions
Intel, Advanced Micro Devices (AMD)	470	464	3674	Electronic & Other Electrical Equipment & Components
Jack in the Box	15	418	5812	Eating and Drinking Places
JP Morgan Chase	181	528	6020	Depository Institutions
Kelly Services	80	433	7363	Business Services
Key Bank	59	494	6020	Depository Institutions
LexisNexis, Sprechman & Associates	39	419	2721	Printing, Publishing and Allied Industries
Lincoln National Corporation (Lincoln Financial)	5	379	6311	Insurance Carriers
McDonald's	10	599	5812	Eating and Drinking Places
Metropolitan Life Insurance Company (MetLife)	146	587	6311	Insurance Carriers
Morgan Stanley	596	370	6211	Security & Commodity Brokers, Dealers, Exchanges & Services
MySpace	24	472	2759	Printing, Publishing and Allied Industries
NASDAQ OMX Group Inc.	21	573	6200	Security & Commodity Brokers, Dealers, Exchanges & Services
Netflix	65	489	7841	Motion Pictures
Nielsen	293	406	8700	Engineering, Accounting, Research, and Management Services
Nissan USA	163	696	3711	Transportation Equipment
Nvidia	44	379	3674	Electronic & Other Electrical Equipment & Components
Oak River Insurance Institute	1	690	6531	Real Estate
Occidental Petroleum Corporation	2	380	1311	Oil and Gas Extraction
PNC Bank	232	605	6020	Depository Institutions
RBC Bank	22	652	6020	Depository Institutions
Regions Bank	49	503	6020	Depository Institutions
Sam's Club	160	675	5331	General Merchandise Stores
SeaChange International	6	616	7372	Business Services
Shell Oil Co.	16	665	2911	Petroleum Refining and Related Industries
Sprint	146	436	4812	Communications
Staples (Staples Business Depot)	247	397	5110	Wholesale Trade - Nondurable Goods
Stephen F. Austin Hotel	72	684	7011	Hotels, Rooming Houses, Camps, and Other Lodging Places
Suddenlink Communications	14	447	4841	Communications
SunTrust Bank	102	501	6020	Depository Institutions
Symantec	169	456	7372	Business Services
Taco Bell	29	370	5812	Eating and Drinking Places
TD Bank	29	437	6020	Depository Institutions
Thomson Reuters	232	588	2741	Printing, Publishing and Allied Industries
Time Warner Cable	72	574	4888	Communications
Tribune Co.	27	440	4833	Communications
TSYS	58	670	7389	Business Services
Union Bank	44	460	6020	Depository Institutions
US Airways	49	461	4512	Transportation by Air
Wells Fargo	438	510	6020	Depository Institutions
Wendy's	37	574	5812	Eating and Drinking Places
Windstream	27	391	4813	Communications
Wyndham Vacation Ownership	45	529	6531	Real Estate

Table C.1 Firms with insider breaches used in this study

Non-breached Firms	Number of reviews analyzed	Number of days of reviews	SIC code	two digit SIC code description
Verizon	659	536	4812	Communications
Starbucks	569	659	5812	Eating and Drinking Places
Pearson PLC	503	445	2731	Printing, Publishing and Allied Industries
Amazon.com	369	550	5961	Miscellaneous Retail
Merck & Co. inc	277	504	2834	Chemicals and Allied Products
Comcast corp	256	512	4841	Communications
Exxon mobil corp	208	442	2911	Petroleum Refining and Related Industries
Allstate	205	440	6331	Insurance Carriers
Marriott international inc	194	514	7011	Hotels, Rooming Houses, Camps, and Other Lodging Places
Discover financial service	192	539	6141	Nondepository Credit Institutions
T-mobile	170	564	4812	Communications
Walt disney company	165	505	4888	Communications
Chevron	159	502	2911	Petroleum Refining and Related Industries
MGM resorts international	126	513	7990	Amusement and Recreation Services
Prudential financial inc	114	630	6311	Insurance Carriers
Ford motor company	102	466	3711	Transportation Equipment
progressive	99	546	6331	Insurance Carriers
ConocoPhillips	90	463	1311	Oil and Gas Extraction
Equifax inc	89	706	7323	Business Services
Westpac banking	84	499	6020	Depository Institutions
RR Donnelley & sons	82	472	2750	Printing, Publishing and Allied Industries
Micron technology	81	601	3674	Electronic & Other Electrical Equipment & Components
Moody's Corp	76	591	7323	Business Services
Viacom inc	75	548	4833	Communications
Volt information sciences inc	72	652	7363	Business Services
Express scripts holding company	65	509	5912	Miscellaneous Retail
Ally financial	57	557	6172	Nondepository Credit Institutions
United continental holdings	56	693	4512	Transportation by Air
Alaska Air group inc	55	529	4512	Transportation by Air
Fitbit inc	52	625	3663	Electronic & Other Electrical Equipment & Components
Tesla inc	48	652	3711	Transportation Equipment
East West Bancorp	46	533	6020	Depository Institutions
Roche Holding AG	40	305	2834	Chemicals and Allied Products
Sonic corp	39	618	5812	Eating and Drinking Places
Tuesday morning corp	36	570	5331	General Merchandise Stores
Cogent communications holdings	33	662	4813	Communications
Bank of Montreal	31	553	6020	Depository Institutions
Sun life financial inc	28	718	6311	Insurance Carriers
Boingo wireless	26	687	4899	Communications
Acco Brands corp	24	722	2780	Printing, Publishing and Allied Industries
Diplomat pharmacy	23	700	5912	Miscellaneous Retail
Navient Corp	20	266	6111	Nondepository Credit Institutions
CME group inc	17	564	6200	Security & Commodity Brokers, Dealers, Exchanges & Services
Meredith corp	16	435	2721	Printing, Publishing and Allied Industries
OSI systems inc	14	562	3844	Measuring, Photographic, Medical, & Optical Goods, & Clocks
Quantenna communications inc	12	429	3674	Electronic & Other Electrical Equipment & Components
Rambus inc	11	615	6794	Holding and Other Investment Offices
Arris group inc	10	297	3663	Electronic & Other Electrical Equipment & Components
Sierra wireless	10	655	3661	Electronic & Other Electrical Equipment & Components
Akamai technologies	9	531	7370	Business Services
Beacon roofing supply	9	567	5030	Wholesale Trade - Durable Goods
Cullen/Frost bank	9	526	6020	Depository Institutions
Cathay general bancorp	5	555	6020	Depository Institutions
Concurrent computer corp	5	437	9995	Nonclassifiable Establishments
Daily journal corp	5	581	7372	Business Services
Intercontinental exchange	5	282	6200	Security & Commodity Brokers, Dealers, Exchanges & Services
Netgear inc	5	535	3576	Industrial and Commercial Machinery and Computer Equipment
Encana corp	4	307	1311	Oil and Gas Extraction
Energen corp	4	447	4924	Electric, Gas and Sanitary Services
Market leader inc (constellation software)	4	683	7372	Business Services
Santander holdings USA inc	4	584	6141	Nondepository Credit Institutions
Ubiquiti networks	4	336	3663	Electronic & Other Electrical Equipment & Components
Cable one inc	3	355	4841	Communications
Cboe global markets inc	3	412	6200	Security & Commodity Brokers, Dealers, Exchanges & Services
Euronet worldwide inc	2	71	6099	Depository Institutions
Beneficial Bancorp	1	1	6035	Depository Institutions
Century bancorp	1	1	6020	Depository Institutions
Envision healthcare corp	1	1	8011	Health Services
First Bancorp	1	1	6020	Depository Institutions
Floor & décor holdings	1	1	5211	Building Materials, Hardware, Garden Supplies & Mobile Homes

Table C.2. Firms with no insider breaches

```
Call:
glm(formula = smalldata_with_anxiety$Breached_or_not ~ smalldata_with_anxiety$rating_overall_median +
    smalldata_with_anxiety$rating_comp_median + smalldata_with_anxiety$Anger_PC +
    smalldata_with_anxiety$Fear_PC + smalldata_with_anxiety$Joy_PC +
    smalldata_with_anxiety$Sadness_PC + smalldata_with_anxiety$Anxiety_PC,
    family = binomial(link = "logit"), data = smalldata_with_anxiety)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.02731  -0.32974   0.03908   0.37793   1.91934

Coefficients:
                                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                                   54.901     16.055   3.420 0.000627 ***
smalldata_with_anxiety$rating_overall_median  -2.510      1.146  -2.191 0.028473 *
smalldata_with_anxiety$rating_comp_median     -3.297      1.380  -2.389 0.016904 *
smalldata_with_anxiety$Anger_PC                9.220      5.832   1.581 0.113851
smalldata_with_anxiety$Fear_PC              -106.190     38.916  -2.729 0.006358 **
smalldata_with_anxiety$Joy_PC                -51.522     18.647  -2.763 0.005727 **
smalldata_with_anxiety$Sadness_PC             -5.668      3.521  -1.610 0.107374
smalldata_with_anxiety$Anxiety_PC             29.060      8.448   3.440 0.000582 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 97.283  on 70  degrees of freedom
Residual deviance: 39.790  on 63  degrees of freedom
AIC: 55.79

Number of Fisher Scoring iterations: 8
```

Figure C.3 Logistic regression result model (1)

```
Call:
glm(formula = smalldata_with_anxiety$Breached_or_not ~ smalldata_with_anxiety$rating_overall_median +
    smalldata_with_anxiety$rating_comp_median + smalldata_with_anxiety$Anger_PC +
    smalldata_with_anxiety$Fear_PC + smalldata_with_anxiety$Joy_PC +
    smalldata_with_anxiety$Sadness_PC, family = binomial(link = "logit"),
    data = smalldata_with_anxiety)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3571  -0.8271   0.4868   0.8014   1.9107

Coefficients:
                                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                                  32.4323     9.7994   3.310 0.000934 ***
smalldata_with_anxiety$rating_overall_median -1.1216     0.6258  -1.792 0.073104 .
smalldata_with_anxiety$rating_comp_median    -1.4396     0.8294  -1.736 0.082618 .
smalldata_with_anxiety$Anger_PC               1.4952     2.9080   0.514 0.607139
smalldata_with_anxiety$Fear_PC              -33.0924    18.9974  -1.742 0.081518 .
smalldata_with_anxiety$Joy_PC               -30.9748    12.8358  -2.413 0.015815 *
smalldata_with_anxiety$Sadness_PC            -4.2168     2.5652  -1.644 0.100205
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 97.283  on 70  degrees of freedom
Residual deviance: 73.429  on 64  degrees of freedom
AIC: 87.429

Number of Fisher Scoring iterations: 6
```

Figure C.4 Logistic regression result model (2)

Chapter 5

General Conclusions

The purpose of this dissertation research is to explore the intersections of data analytics and cybersecurity. The objective of the three dissertation essays is to learn more about cybersecurity analytics by applying data analytics techniques. The first essay attempts to understand the intellectual structure of business analytics research carried out by information systems researchers based on the articles published in the senior scholar's basket of eight journals within the last two decades. One of the aims of this dissertation essay is to identify the underlying themes latent in the corpus of BA IS research. From this first essay, I was able to identify various sub-disciplines within the IS business analytics community, which were underrepresented in the elite IS journals. The study of the security breach is one such sub-discipline. The second and the third essays in this project attempt to fill these intellectual and practical gaps by carrying out research in the domain of cyber security analytics.

From the tools, techniques, and methodologies perspective, the first dissertation essay employs techniques of inter-citation counts, topic modeling, and social network analysis via Exponential Random Graph Modeling (ERGM). The second essay uses alternative techniques of topic modeling—cosine similarity and social network analysis—based on Quadratic Assignment Procedure (QAP). Hence, the first and the second essays overlap in their methodologies in that both are based on usage of certain text and social network analysis techniques. However, the second essay explores the unique phenomenon of breaching together, where we have not considered the types of breaches. The final dissertation essay uses matched sample technique, techniques of psycholinguistics, and logistic regression to gain meaningful insights towards understanding of security breach committed intentionally by insiders. Hence, the second and third essays are also related, as both use breaches as the dependent variable. However, the second essay studies breaches from a dyadic perspective whereas third essay considers a single category of security breach.

Data analytics and cybersecurity are evolving fields that are important both for the academic researchers and practitioners. These topics are highly interdisciplinary and concern researchers from domains as varied as computer science, information systems, information security, psychology, linguistics, and accounting and management, among others. The diverse tools, techniques, methodologies, and theories used in this dissertation make that diverse relevance evident. Hence, it is important for a domain expert working in cybersecurity to expand her knowledge by collaborating with researchers from other domains or by consulting research conducted in other disciplines to gain better understanding of security in cyberspace. The blurring of disciplinary boundaries is inevitable in the information age, and it is impossible for a domain expert to work one field alone. This dissertation makes that reality clear to both academic researchers and industry practitioners.

References

Abbasi, A., Albrecht, C., Vance, A., & Hansen, J. (2012). Metafraud: a meta-learning framework for detecting financial fraud. *Mis Quarterly*, *36*(4), 1293–1327.

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2), 3.

Abbasi, A., Zahedi, F. "Mariam," Zeng, D., Chen, Y., Chen, H., & Nunamaker, J. F. (2015). Enhancing Predictive Analytics for Anti-Phishing by Exploiting Website Genre Information. *Journal of Management Information Systems*, *31*(4), 109–157. https://doi.org/10.1080/07421222.2014.1001260

Acquisti, A., Friedman, A., & Telang, R. (2006). Is there a cost to privacy breaches? An event study. *ICIS 2006 Proceedings*, 94.

Agarwal, R. (2016). Editorial—On the Intellectual Structure and Evolution of *ISR*. *Information Systems Research*, *27*(3), 471–477. https://doi.org/10.1287/isre.2016.0670

Agarwal, R., & Dhar, V. (2014). **Editorial** —Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. *Information Systems Research*, *25*(3), 443–448. https://doi.org/10.1287/isre.2014.0546

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In *Action control* (pp. 11–39). Springer.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211.

Aldag, R. J., & Brief, A. P. (1979). *Task design and employee motivation*. Scott Foresman.

Aldous, D. J. (1985). Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XIII—1983* (pp. 1–198). Springer.

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, *23*(4), 589. https://doi.org/10.2307/2978933

Anagnostopoulos, A., Kumar, R., & Mahdian, M. (2008). Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 7–15). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1401897

Anderson, R., & Moore, T. (2009). Information security: where computer science, economics and psychology meet. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *367*(1898), 2717–2727. https://doi.org/10.1098/rsta.2009.0027

Anderson, Ross, & Moore, T. (2007). Information security economics–and beyond. In *Annual International Cryptology Conference* (pp. 68–91). Springer.

Angst, C. M., Block, E. S., D'arcy, J., & Kelley, K. (2017). When do IT security investments matter? Accounting for the influence of institutional factors in the context of healthcare data breaches. *MIS Quarterly*, *41*(3), 893–916.

Arnott, D., & Pervan, G. (2005). A critical analysis of decision support systems research. *Journal of Information Technology*, *20*(2), 67–87. https://doi.org/10.1057/palgrave.jit.2000035

Arnott, D., & Pervan, G. (2012). Design science in decision support systems research: An assessment using the Hevner, March, Park, and Ram Guidelines. *Journal of the Association for Information Systems*, *13*(11), 923.

Arnott, D., & Pervan, G. (2014). A critical analysis of decision support systems research revisited: the rise of design science. *Journal of Information Technology*, *29*(4), 269–293.

Ashbaugh-skaife, H., & Lafond, R. (2006). The Effect of Internal Control Deficiencies on Firm Risk and Cost of Equity Capital" Working paper.

Audzeyeva, A., & Hudson, R. (2016). How to get the most from a business intelligence application during the post implementation phase? Deep structure transformation at a UK retail bank. *European Journal of Information Systems*, *25*(1), 29–46.

Avery, A., & Ranganathan, C. (2016). Financial Performance Impacts of Information Security Breaches.

Baer, W. S., & Parkinson, A. (2007). Cyberinsurance in it security management. *IEEE Security & Privacy*, *5*(3).

Baker, J., Jones, D. R., & Burkman, J. (2009). Using visual representations of data to enhance sensemaking in data exploration tasks. *Journal of the Association for Information Systems*, *10*(7), 533.

Balijepally, V., & Nerur, S. (2015). Understanding the Structure of Agile Software Development Using Text Analytics: A Preliminary Analysis.

Bandura, A. (1978). Social learning theory of aggression. *Journal of Communication*, *28*(3), 12–29.

Bardhan, I., Krishnan, V., & Lin, S. (2013). Research Note—Business Value of Information Technology: Testing the Interaction Effect of IT and R&D on Tobin's Q. *Information Systems Research*, *24*(4), 1147–1161.

Bardhan, I., Oh, J. (Cath), Zheng, Z. (Eric), & Kirksey, K. (2015). Predictive Analytics for Readmission of Patients with Congestive Heart Failure. *Information Systems Research*, *26*(1), 19–39. https://doi.org/10.1287/isre.2014.0553

Baroudi, J. J. (1985). The Impact of Role Variables on IS Personnel Work Attitudes and Intentions. *MIS Quarterly*, *9*(4), 341. https://doi.org/10.2307/249234

Ben Letaifa, S., & Rabeau, Y. (2013). Too close to collaborate? How geographic proximity could impede entrepreneurship and innovation. *Journal of Business Research*, *66*(10), 2071–2078. https://doi.org/10.1016/j.jbusres.2013.02.033

Bernstein, A., Bienstock, D., Hay, D., Uzunoglu, M., & Zussman, G. (2011). Power Grid Vulnerability to Geographically Correlated Failures.

Berry, M. W., & Castellanos, M. (2004). Survey of text mining. *Computing Reviews*, *45*(9), 548.

Bharadwaj, A. S. (2000). A Resource-Based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation. *MIS Quarterly*, *24*(1), 169. https://doi.org/10.2307/3250983

Biener, C., Eling, M., & Wirfs, J. H. (2015). Insurability of cyber risk: An empirical analysis. *The Geneva Papers on Risk and Insurance-Issues and Practice*, *40*(1), 131–158.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Böhme, R., & Kataria, G. (2006). Models and Measures for Correlation in Cyber-Insurance. In *WEIS*.

Böhme, R., & Schwartz, G. (2010). Modeling Cyber-Insurance: Towards a Unifying Framework. In *WEIS*.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, *323*(5916), 892–895. https://doi.org/10.1126/science.1165821

Borgman, C. L. (1990). Scholarly communication and bibliometrics.

Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies*, *39*(1), 61–74. https://doi.org/10.1080/0034340052000320887

Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, *61*(12), 2389–2404. https://doi.org/10.1002/asi.21419

Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, *64*(3), 351–374.
boyd, danah m., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x
Breslow, N. E., Day, N. E., & others. (1980). *Statistical methods in cancer research. Vol. 1. The analysis of case-control studies.* (Vol. 1). Distributed for IARC by WHO, Geneva, Switzerland.
Bulgurcu, Cavusoglu, & Benbasat. (2010). Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness. *MIS Quarterly*, *34*(3), 523. https://doi.org/10.2307/25750690
Buzan, B. (2008). *People, States & Fear: An agenda for international security studies in the post-cold war era*. Ecpr Press.
Buzan, B., & Waever, O. (2003). *Regions and powers: the structure of international security* (Vol. 91). Cambridge University Press.
Cachin, C., & Schunter, M. (2011). A cloud you can trust. *IEEE Spectrum*, *48*(12), 28–51.
Campbell, K., Gordon, L. A., Loeb, M. P., & Zhou, L. (2003). The economic cost of publicly announced information security breaches: empirical evidence from the stock market. *Journal of Computer Security*, *11*(3), 431–448.
Charvet, F. F., Cooper, M. C., & Gardner, J. T. (2008). The intellectual structure of supply chain management: A bibliometric approach. *Journal of Business Logistics*, *29*(1), 47–73.
Chau, M., & Xu, J. (2012). Business intelligence in blogs: Understanding consumer interactions and communities. *MIS Quarterly*, *36*(4), 1189–1216.
Chellappa, R. K., Sambamurthy, V., & Saraf, N. (2010). Competing in Crowded Markets: Multimarket Contact and the Nature of Competition in the Enterprise Systems Software Industry. *Information Systems Research*, *21*(3), 614–630. https://doi.org/10.1287/isre.1100.0300
Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management. *Journal of Management Information Systems*, *32*(4), 4–39. https://doi.org/10.1080/07421222.2015.1138364
Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, *36*(4), 1165–1188.
Chen, Kataria, & Krishnan. (2011). Correlated Failures, Diversification, and Information Security Risk Management. *MIS Quarterly*, *35*(2), 397. https://doi.org/10.2307/23044049
Choudhary, V., & Vithayathil, J. (2013). The Impact of Cloud Computing: Should the IT Department Be Organized as a Cost Center or a Profit Center? *Journal of Management Information Systems*, *30*(2), 67–100. https://doi.org/10.2753/MIS0742-1222300203
Chuang, J., Manning, C. D., & Heer, J. (2012). Termite: Visualization techniques for assessing textual topic models. In *Proceedings of the international working conference on advanced visual interfaces* (pp. 74–77). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2254572
Chung, W., Chen, H., & Nunamaker Jr, J. F. (2005). A visual framework for knowledge discovery on the Web: An empirical study of business intelligence exploration. *Journal of Management Information Systems*, *21*(4), 57–84.
Clark, T. D., Jones, M. C., & Armstrong, C. P. (2007). The dynamic structure of management support systems: theory development, research focus, and direction. *Mis Quarterly*, *31*(3), 579–615.
Clarke, R. (2016). Big data, big risks: Big data, big risks. *Information Systems Journal*, *26*(1), 77–90. https://doi.org/10.1111/isj.12088
Clarke, R. V. (1980). Situational crime prevention: Theory and practice. *Brit. J. Criminology*, *20*, 136.

Coelho, C. G., Falótico, T., Izar, P., Mannu, M., Resende, B. D., Siqueira, J. O., & Ottoni, E. B. (2015). Social learning strategies for nut-cracking by tufted capuchin monkeys (Sapajus spp.). *Animal Cognition*, *18*(4), 911–919. https://doi.org/10.1007/s10071-015-0861-5

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*(1), 128. https://doi.org/10.2307/2393553

Coleman, J. S. (2000). Social capital in the creation of human capital. In *Knowledge and social capital* (pp. 17–41). Elsevier.

Coltman, T., Devinney, T. M., & Midgley, D. F. (2011). Customer relationship management and firm performance. *Journal of Information Technology*, *26*(3), 205–219. https://doi.org/10.1057/jit.2010.39

Constant, D., Sproull, L., & Kiesler, S. (1996). The kindness of strangers: The usefulness of electronic weak ties for technical advice. *Organization Science*, *7*(2), 119–135.

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: big data and the changing context of strategy. *Journal of Information Technology*, *30*(1), 44–57.

Cooper, A. (2012). What is analytics? Definition and essential characteristics. *CETIS Analytics Series*, *1*(5), 1–10.

Cron, W. L., & Sobol, M. G. (1983). The relationship between computerization and performance: a strategy for maximizing the economic benefits of computerization. *Information & Management*, *6*(3), 171–181.

Culnan, M. J. (1986). The intellectual development of management information systems, 1972–1982: A co-citation analysis. *Management Science*, *32*(2), 156–172.

Culnan, M. J. (1987). Mapping the Intellectual Structure of MIS, 1980-1985: A Co-Citation Analysis. *MIS Quarterly*, *11*(3), 341. https://doi.org/10.2307/248680

Culnan, M. J., O'Reilly, C. A., & Chatman, J. A. (1990). Intellectual structure of research in organizational behavior, 1972-1984: a cocitation analysis. *Journal of the American Society for Information Science*, *41*(6), 453.

Culnan, & Williams. (2009). How Ethics Can Enhance Organizational Privacy: Lessons from the Choicepoint and TJX Data Breaches. *MIS Quarterly*, *33*(4), 673. https://doi.org/10.2307/20650322

Damanpour, F. (1992). Organizational size and innovation. *Organization Studies*, *13*(3), 375–402.

Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., & Lee, L. (2009). How opinions are received by online communities: A case study on Amazon.com helpfulness votes. *ArXiv:0906.3741 [Physics]*. Retrieved from http://arxiv.org/abs/0906.3741

D'Arcy, J., Hovav, A., & Galletta, D. (2009). User Awareness of Security Countermeasures and Its Impact on Information Systems Misuse: A Deterrence Approach. *Information Systems Research*, *20*(1), 79–98. https://doi.org/10.1287/isre.1070.0160

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Press.

Davis, G. B. (2006). Information systems as an academic discipline. In *The Past and Future of Information Systems: 1976–2006 and Beyond* (pp. 11–25). Springer.

Deeds, D. L. (2001). The role of R&D intensity, technical development and absorptive capacity in creating entrepreneurial wealth in high technology start-ups. *Journal of Engineering and Technology Management*, *18*(1), 29–47. https://doi.org/10.1016/S0923-4748(00)00032-1

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (n.d.). Indexing by Latent Semantic Analysis.

DeKay, S. H. (2013). Peering Through Glassdoor. Com What Social Media Can Tell Us About Employee Satisfaction and. In *CONFERENCE ON CORPORATE COMMUNICATION 2013* (Vol. 45).
Dekker, D., Krackhardt, D., & Snijders, T. (2003). Multicollinearity robust QAP for multiple regression. In *1st annual conference of the North American Association for Computational Social and Organizational Science* (pp. 22–25).
Dekker, D., Krackhardt, D., & Snijders, T. A. B. (2007). Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions. *Psychometrika*, *72*(4), 563–581. https://doi.org/10.1007/s11336-007-9016-1
Demerouti, E., Bakker, A. B., & Bulters, A. J. (2004). The loss spiral of work pressure, work–home interference and exhaustion: Reciprocal relations in a three-wave study. *Journal of Vocational Behavior*, *64*(1), 131–149. https://doi.org/10.1016/S0001-8791(03)00030-7
Denecke, K., & Deng, Y. (2015). Sentiment analysis in medical settings: New opportunities and challenges. *Artificial Intelligence in Medicine*, *64*(1), 17–27. https://doi.org/10.1016/j.artmed.2015.03.006
Deng, X., & Chi, L. (2012). Understanding Postadoptive Behaviors in Information Systems Use: A Longitudinal Analysis of System Use Problems in the Business Intelligence Context. *Journal of Management Information Systems*, *29*(3), 291–326. https://doi.org/10.2753/MIS0742-1222290309
Detmar W. Straub, & Jr. (1990). Effective IS Security: An Empirical Study. *Information Systems Research*, *1*(3), 255–276.
Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, *56*(12), 64–73.
Ding, Y. (2011). Community detection: Topological vs. topical. *Journal of Informetrics*, *5*(4), 498–514.
Dong, Y., Huang, X., Sinha, K. K., & Xu, K. (2014). Collaborative Demand Forecasting: Toward the Design of an Exception-Based Forecasting Mechanism. *Journal of Management Information Systems*, *31*(2), 245–284. https://doi.org/10.2753/MIS0742-1222310209
Doyle, J., Ge, W., & McVay, S. (2007). Determinants of weaknesses in internal control over financial reporting. *Journal of Accounting and Economics*, *44*(1–2), 193–223. https://doi.org/10.1016/j.jacceco.2006.10.003
Durkin, K. F., Wolfe, T. W., & Clark, G. (1999). Social bond theory and binge drinking among college students: A multivariate analysis. *College Student Journal*, *33*, 450–462.
Duttweiler, P. C. (1986). Educational excellence and motivating teachers. *The Clearing House*, *59*(8), 371–374.
Ellison, G., & Glaeser, E. L. (1997). Geographic concentration in US manufacturing industries: a dartboard approach. *Journal of Political Economy*, *105*(5), 889–927.
Eom, S. B. (1996). Mapping the intellectual structure of research in decision support systems through author cocitation analysis (1971–1993). *Decision Support Systems*, *16*(4), 315–338.
Ernest Chang, S., & Lin, C. (2007). Exploring organizational culture for information security management. *Industrial Management & Data Systems*, *107*(3), 438–458. https://doi.org/10.1108/02635570710734316
Ethiraj, S. K., Kale, P., Krishnan, M. S., & Singh, J. V. (2005). Where do capabilities come from and how do they matter? A study in the software services industry. *Strategic Management Journal*, *26*(1), 25–45. https://doi.org/10.1002/smj.433
Fan, W., Gordon, M. D., Pathak, P., & PATHAK, P. (2005). Genetic programming-based discovery of ranking functions for effective web search. *Journal of Management Information Systems*, *21*(4), 37–56.
Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, *49*(9), 76–82.

Faraj, S., & Johnson, S. L. (2011). Network Exchange Patterns in Online Communities. *Organization Science*, *22*(6), 1464–1480. https://doi.org/10.1287/orsc.1100.0600

Faulkner, B. (2007). Hacking into data breach notification laws. *Fla. L. Rev.*, *59*, 1097.

Fayard, A.-L., Gkeredakis, E., & Levina, N. (2016). Framing Innovation Opportunities While Staying Committed to an Organizational Epistemic Stance. *Information Systems Research*, *27*(2), 302–323. https://doi.org/10.1287/isre.2016.0623

Figueiredo, M. A. T., & Jain, A. K. (2002). Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(3), 381–396.

Fodor, J., Bever, A., Garrett, T., & others. (1974). The psychology of language: An introduction to psycholinguistics and generative grammar.

Folkman, S., & Lazarus, R. S. (1988). Coping as a mediator of emotion. *Journal of Personality and Social Psychology*, *54*(3), 466.

Folkman, S., Lazarus, R. S., Dunkel-Schetter, C., DeLongis, A., & Gruen, R. J. (1986). Dynamics of a stressful encounter: Cognitive appraisal, coping, and encounter outcomes. *Journal of Personality and Social Psychology*, *50*(5), 992–1003. https://doi.org/10.1037/0022-3514.50.5.992

Frans A. J. Van den Bosch, Henk W. Volberda, & Michiel de Boer. (n.d.). Coevolution of Firm Absorptive Capacity and Knowledge Environment: Organizational Forms and Combinative Capabilities. *Organization Science*, *10*(5,). Retrieved from http://www.jstor.org/stable/2640317

Freeman, L. C. (1996). Some antecedents of social network analysis. *Connections*, *19*(1), 39–42.

Freeman, L. C. (2004). *The development of social network analysis: a study in the sociology of science*. Vancouver, BC : North Charleston, S.C: Empirical Press ; BookSurge.

Gatzlaff, K. M., & McCullough, K. A. (2010). The Effect of Data Breaches on Shareholder Wealth. *Risk Management and Insurance Review*, *13*(1), 61–83. https://doi.org/10.1111/j.1540-6296.2010.01178.x

Ge, W., & McVay, S. (2005). The Disclosure of Material Weaknesses in Internal Control after the Sarbanes-Oxley Act. *Accounting Horizons*, *19*(3), 137–158. https://doi.org/10.2308/acch.2005.19.3.137

Gholami, R., Watson, R. T., Molla, A., Hasan, H., & Bjørn-Andersen, N. (2016). Information systems solutions for environmental sustainability: How can we do more? *Journal of the Association for Information Systems*, *17*(8), 521.

Ghoshal, A., Menon, S., & Sarkar, S. (2015). Recommendations Using Information from Multiple Association Rules: A Probabilistic Approach. *Information Systems Research*, *26*(3), 532–551. https://doi.org/10.1287/isre.2015.0583

Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. *Icwsm*, *7*(21), 219–222.

Golub, G. H., & Reinsch, C. (1970). Singular value decomposition and least squares solutions. *Numerische Mathematik*, *14*(5), 403–420.

Goodreau, S. M. (2007). Advances in exponential random graph (p*) models applied to a large social network. *Social Networks*, *29*(2), 231–248.

Goodreau, S. M., Handcock, M. S., Hunter, D. R., Butts, C. T., & Morris, M. (2008). A statnet Tutorial. *Journal of Statistical Software*, *24*(9), 1.

Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography*, *46*(1), 103–125.

Gordon, L. A., & Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC)*, *5*(4), 438–457.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime.* Stanford University Press.

Gou, L., Zhou, M. X., & Yang, H. (2014). KnowMe and ShareMe: understanding automatically discovered personality traits from social media and user sharing preferences (pp. 955–964). ACM Press. https://doi.org/10.1145/2556288.2557398

Granovetter, M. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, *1*, 201. https://doi.org/10.2307/202051

Granovetter, M. S. (1977). The strength of weak ties. In *Social networks* (pp. 347–367). Elsevier.

Greenhaus, J. H., & Beutell, N. J. (1985). Sources of Conflict between Work and Family Roles. *The Academy of Management Review*, *10*(1), 76. https://doi.org/10.2307/258214

Greenwald, A., Kannan, K., & Krishnan, R. (2010). On Evaluating Information Revelation Policies in Procurement Auctions: A Markov Decision Process Approach. *Information Systems Research*, *21*(1), 15–36. https://doi.org/10.1287/isre.1080.0168

Gritzalis, D. (2014). *Holistic Information Security: Human Factor and Behavior Prediction using Social Media*. Citeseer.

Grover, V., Cheon, M. J., & Teng, J. T. (1994). A descriptive study on the outsourcing of information systems functions. *Information & Management*, *27*(1), 33–44.

Grover, V., Gokhale, R., Lim, J., Coffey, J., & Ayyagari, R. (2006). A citation analysis of the evolution and state of information systems within a constellation of reference disciplines. *Journal of the Association for Information Systems*, *7*(5), 13.

Grus, J. (2015). *Data science from scratch: first principles with python*. O'Reilly Media, Inc.

Gulati, R. (1998). Alliances and networks. *Strategic Management Journal*, *19*(4), 293–317.

Gupta, V., Lehal, G. S., & others. (2009). A survey of text mining techniques and applications. *Journal of Emerging Technologies in Web Intelligence*, *1*(1), 60–76.

Habjan, A., Andriopoulos, C., & Gotsi, M. (2014). The role of GPS-enabled information in transforming operational decision making: an exploratory study. *European Journal of Information Systems*, *23*(4), 481–502. https://doi.org/10.1057/ejis.2014.2

Haimann, T. (1973). *Supervisory management for health care institutions*. Catholic Hospital Association.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L., & others. (2006). *Multivariate data analysis (Vol. 6)*. Upper Saddle River, NJ: Pearson Prentice Hall.

Hamner, W. C., & Organ, D. W. (1978). *Organizational behavior: An applied psychological approach*. Business Publications.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, *24*(1), 1548.

Harris, J. K. (2013a). *An introduction to exponential random graph modeling* (Vol. 173). Sage Publications.

Harris, J. K. (2013b). *An introduction to exponential random graph modeling* (Vol. 173). Sage Publications.

Heath, D., & Sudderth, W. (1976). De Finetti's theorem on exchangeable variables. *The American Statistician*, *30*(4), 188–189.

Heath, N. M., Hall, B. J., Russ, E. U., Canetti, D., & Hobfoll, S. E. (2012). Reciprocal relationships between resource loss and psychological distress following exposure to political violence: an empirical investigation of COR theory's loss spirals. *Anxiety, Stress & Coping*, *25*(6), 679–695. https://doi.org/10.1080/10615806.2011.628988

Heckman, P. E., & Meyers, G. G. (1983). The calculation of aggregate loss distributions from claim severity and claim count distributions. In *Proceedings of the Casualty Actuarial Society* (Vol. 70).

Hersey, P., Blanchard, K. H., & Johnson, D. E. (2007). *Management of organizational behavior* (Vol. 9). Prentice hall Upper Saddle River, NJ.

Herzberg, F. (1976). *The managerial choice: To be efficient and to be human*. Irwin Professional Publishing.

Herzberg, F. I. (1966). Work and the nature of man.

Herzberg, F., Mausner, B. S., & Snyderman, B. (1978). B.(1959). The motivation to work. *New York: John Willey*.

Hirschi, T. (1969). 1969 Causes of Delinquency Berkeley, Calif.: University of California Press.

Hoberg, G., & Phillips, G. (2010). Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *Review of Financial Studies*, *23*(10), 3773–3811. https://doi.org/10.1093/rfs/hhq053

Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, *124*(5), 1423–1465.

Hobfoll, S. E. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist*, *44*(3), 513.

Hofmann, T. (1999). Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence* (pp. 289–296). Morgan Kaufmann Publishers Inc.

Hofmann, T., Puzicha, J., & Jordan, M. I. (1999). Learning from dyadic data. In *Advances in neural information processing systems* (pp. 466–472).

Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. *Decision Support Systems*, *64*, 130–141. https://doi.org/10.1016/j.dss.2014.05.013

Hossack, I., Pollard, J. H., & Zehnwirth, B. (1999). *Introductory statistics with applications in general insurance*. Cambridge University Press.

Hotho, A., Nürnberger, A., & Paa\s s, G. (2005). A brief survey of text mining. In *Ldv Forum* (Vol. 20, pp. 19–62). Retrieved from http://www.kde.cs.uni-kassel.de/hotho/pub/2005/hotho05TextMining.pdf

Hu, D., Zhao, J. L., Hua, Z., & Wong, M. C. (2012). Network-based modeling and analysis of systemic risk in banking systems. *MIS Quarterly*, *36*(4), 1269–1291.

Hu, N., Liu, L., & Zhang, J. J. (2008). Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Information Technology and Management*, *9*(3), 201–214. https://doi.org/10.1007/s10799-008-0041-2

Hu, N., Pavlou, P. A., & Zhang, J. (2006). Can online reviews reveal a product’s true quality?: empirical findings and analytical modeling of Online word-of-mouth communication. In *Proceedings of the 7th ACM conference on Electronic commerce* (pp. 324–330). ACM.

Huang, C. D., Hu, Q., & Behara, R. S. (2006). Economics of Information Security Investment in the Case of Simultaneous Attacks. In *WEIS*. Citeseer.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008a). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, *24*(3), nihpa54860.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008b). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, *24*(3), nihpa54860.

Igbaria, M., & Guimaraes, T. (1993). Antecedents and consequences of job satisfaction among information center employees. *Journal of Management Information Systems*, *9*(4), 145–174.

Ingersoll, G. L., Olsan, T., Drew-Cates, J., DeVinney, B. C., & Davies, J. (2002). Nurses’ job satisfaction, organizational commitment, and career intent. *Journal of Nursing Administration*, *32*(5), 250–263.

Inkpen, A. C., & Tsang, E. W. (2005). Social capital, networks, and knowledge transfer. *Academy of Management Review*, *30*(1), 146–165.

Jabr, W., Mookerjee, R., Tan, Y., & Mookerjee, V. (2013). Leveraging Philanthropic Behavior for Customer Support: The case of user support forums.

Jarneving, B. (2005). A comparison of two bibliometric methods for mapping of the research front. *Scientometrics*, *65*(2), 245–263.

Ji, Y., Rozenbaum, O., & Welch, K. T. (2017). Corporate culture and financial reporting risk: Looking through the Glassdoor.

Jindal, N., & Liu, B. (2008). Opinion spam and analysis. In *Proceedings of the 2008 International Conference on Web Search and Data Mining* (pp. 219–230). ACM.

Jo, Y., & Oh, A. H. (2011). Aspect and sentiment unification model for online review analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 815–824). ACM.

Kayande, U., De Bruyn, A., Lilien, G. L., Rangaswamy, A., & van Bruggen, G. H. (2009). How Incorporating Feedback Mechanisms in a DSS Affects DSS Evaluations. *Information Systems Research*, *20*(4), 527–546. https://doi.org/10.1287/isre.1080.0198

Kim, G., Shin, B., Kim, K. K., & Lee, H. G. (2011). IT capabilities, process-oriented dynamic capabilities, and firm financial performance. *Journal of the Association for Information Systems*, *12*(7), 487.

Kim, H. M., Fox, M. S., & Sengupta, A. (2007). How to build enterprise data models to achieve compliance to standards or regulatory requirements (and share data). *Journal of the Association for Information Systems*, *8*(2), 105.

Ko, M., & Dorantes, C. (2006). The impact of information security breaches on financial performance of the breached firms: an empirical investigation. *Journal of Information Technology Management*, *17*(2), 13–22.

Koh, S. C. L., Gunasekaran, A., & Goodman, T. (2011). Drivers, barriers and critical success factors for ERPII implementation in supply chains: A critical analysis. *The Journal of Strategic Information Systems*, *20*(4), 385–402. https://doi.org/10.1016/j.jsis.2011.07.001

Kosala, R., & Blockeel, H. (2000). Web mining research: A survey. *ACM Sigkdd Explorations Newsletter*, *2*(1), 1–15.

Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, *10*(4), 359–381.

Krackhardt, D. (1992). A caveat on the use of the quadratic assignment procedure. *Journal of Quantitative Anthropology*, *3*(4), 279–296.

Kraemer, S., Carayon, P., & Clem, J. (2009). Human and organizational factors in computer and information security: Pathways to vulnerabilities. *Computers & Security*, *28*(7), 509–520. https://doi.org/10.1016/j.cose.2009.04.006

Kunreuther, H., & Heal, G. (2003). Interdependent security. *Journal of Risk and Uncertainty*, *26*(2–3), 231–249.

Kwon, J., & Johnson, M. E. (2014). PROACTIVE VS. REACTIVE SECURITY INVESTMENTS IN THE HEALTHCARE SECTOR. *Manage Inf Syst Q*, *38*(2), 451–471.

Kwon, J., & Johnson, M. E. (2015). The Market Effect of Healthcare Security: Do Patients Care about Data Breaches? In *WEIS*.

Labianca, G., & Brass, D. J. (2006). Exploring the social ledger: Negative relationships and negative asymmetry in social networks in organizations. *Academy of Management Review*, *31*(3), 596–614.

Lacity, M. C., Solomon, S., Yan, A., & Willcocks, L. P. (2011). Business process outsourcing studies: a critical review and research directions. *Journal of Information Technology*, *26*(4), 221–258. https://doi.org/10.1057/jit.2011.25

Lancaster, T., & Imbens, G. (1991). *Choice-based sampling: Inference and optimality*. University of Bristol, Department of Economics.

Latora, V., & Marchiori, M. (2001). Efficient Behavior of Small-World Networks. *Physical Review Letters*, *87*(19). https://doi.org/10.1103/PhysRevLett.87.198701

Lau, R. Y., Liao, S. S., Wong, K.-F., & Chiu, D. K. (2012). Web 2.0 environmental scanning and adaptive decision support for business mergers and acquisitions. *MIS Quarterly*, *36*(4), 1239–1268.

Lavie, D. (2007). Alliance portfolios and firm performance: A study of value creation and appropriation in the U.S. software industry. *Strategic Management Journal*, *28*(12), 1187–1212. https://doi.org/10.1002/smj.637

Lee, J., & Kang, J. (2017). A Study on Job Satisfaction Factors in Retention and Turnover Groups using Dominance Analysis and LDA Topic Modeling with Employee Reviews on Glassdoor. com.

Leidner, & Kayworth. (2006). Review: A Review of Culture in Information Systems Research: Toward a Theory of Information Technology Culture Conflict. *MIS Quarterly*, *30*(2), 357. https://doi.org/10.2307/25148735

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition & Emotion*, *14*(4), 473–493. https://doi.org/10.1080/026999300402763

Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010). Predicting positive and negative links in online social networks. In *Proceedings of the 19th international conference on World wide web* (pp. 641–650). ACM.

Li, X.-B., & Sarkar, S. (2014). Digression and value concatenation to enable privacy-preserving regression. *MIS Quarterly: Management Information Systems*, *38*(3), 679.

Li, Xin, Chen, H., Zhang, Z., Li, J., & Nunamaker, J. F. (2009). Managing Knowledge in Light of Its Evolution Process: An Empirical Study on Citation Network-Based Patent Classification. *Journal of Management Information Systems*, *26*(1), 129–154. https://doi.org/10.2753/MIS0742-1222260106

Li, Xixi, Hsieh, J. J. P.-A., & Rai, A. (2013). Motivational Differences Across Post-Acceptance Information System Usage Behaviors: An Investigation in the Business Intelligence Systems Context. *Information Systems Research*, *24*(3), 659–682. https://doi.org/10.1287/isre.1120.0456

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167.

Locke, E. A. (1976). The nature and causes of job satisfaction. Handbook of industrial and organizational psichology. *The Nature and Causes of Job Satisfaction: Handbook of Industrial and Organizational Psichology*.

Loebbecke, C., & Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *The Journal of Strategic Information Systems*, *24*(3), 149–157. https://doi.org/10.1016/j.jsis.2015.08.002

Loh, L., & Venkatraman, N. (1992). Determinants of information technology outsourcing: a cross-sectional analysis. *Journal of Management Information Systems*, *9*(1), 7–24.

Luftman, J., Derksen, B., Dwivedi, R., Santana, M., Zadeh, H. S., & Rigoni, E. (2015). Influential IT management trends: an international study. *Journal of Information Technology*, *30*(3), 293–305.

Luftman, J., & Zadeh, H. S. (2011). Key information technology and management issues 2010–11: an international study. *Journal of Information Technology*, *26*(3), 193–204. https://doi.org/10.1057/jit.2011.3

Luftman, J., Zadeh, H. S., Derksen, B., Santana, M., Rigoni, E. H., & Huang, Z. (2013). Key information technology and management issues 2012–2013: an international study. *Journal of Information Technology*, *28*(4), 354–366. https://doi.org/10.1057/jit.2013.22

Luftman, J., Zadeh, H. S., Derksen, B., Santana, M., Rigoni, E. H., & Huang, Z. D. (2012). Key information technology and management issues 2011–2012: an international study. *Journal of Information Technology*, *27*(3), 198–212.

Luo, N., Zhou, Y., & Shon, J. (2016). Employee Satisfaction and Corporate Performance: Mining Employee Reviews on Glassdoor. com.
Lusher, D., Koskinen, J., & Robins, G. (2012). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge University Press.
Mahmood, T., & Afzal, U. (2013). Security analytics: Big data analytics for cybersecurity: A review of trends, techniques and tools. In *Information assurance (ncia), 2013 2nd national conference on* (pp. 129–134). IEEE.
Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370.
Massaro, D. W. (2014). *Understanding language: An information-processing analysis of speech perception, reading, and psycholinguistics*. Academic Press.
Mathew L. A. Hayward. (2002). When Do Firms Learn from Their Acquisition Experience? Evidence from 1990-1995. *Strategic Management Journal*, *23*(1), 21–39.
Maurer, I., Bartsch, V., & Ebers, M. (2011). The Value of Intra-organizational Social Capital: How it Fosters Knowledge Transfer, Innovation Performance, and Growth. *Organization Studies*, *32*(2), 157–185. https://doi.org/10.1177/0170840610394301
McCain, K. W. (1990). Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science*, *41*(6), 433.
McCain, K. W. (1991). Mapping economics through the journal literature: An experiment in journal cocitation analysis. *Journal of the American Society for Information Science*, *42*(4), 290.
McClelland, D. C. (1967). *Achieving society* (Vol. 92051). Simon and Schuster.
McGregor, D. (1960). Theory X and theory Y. *Organization Theory*, *358*, 374.
McLachlan, G., & Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.
McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 415–444.
Mela, C. F., Roos, J., & Deng, Y. (2013). **Invited Paper** —A Keyword History of Marketing Science. *Marketing Science*, *32*(1), 8–18. https://doi.org/10.1287/mksc.1120.0764
Meyer, G., Adomavicius, G., Johnson, P. E., Elidrisi, M., Rush, W. A., Sperl-Hillen, J. M., & O'Connor, P. J. (2014). A Machine Learning Approach to Improving Dynamic Decision Making. *Information Systems Research*, *25*(2), 239–263. https://doi.org/10.1287/isre.2014.0513
Meyers, G., & Schenker, N. (1983). Parameter Uncertainty in the Collective Risk Model. *PCAS LXX*, 111.
Moen, R., & Norman, C. (2006). *Evolution of the PDCA cycle*.
Mookerjee, V. S., & Mannino, M. V. (1997). Redesigning case retrieval to reduce information acquisition costs. *Information Systems Research*, *8*(1), 51–68.
Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal Theories of Emotion: State of the Art and Future Development. *Emotion Review*, *5*(2), 119–124. https://doi.org/10.1177/1754073912468165
Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of exponential-family random graph models: terms and computational aspects. *Journal of Statistical Software*, *24*(4), 1548.
Mowery, D. C., Oxley, J. E., & Silverman, B. S. (1996). Strategic alliances and interfirm knowledge transfer. *Strategic Management Journal*, *17*(S2), 77–91.
Mowery, D. C., Oxley, J. E., & Silverman, B. S. (1998). Technological overlap and interfirm cooperation: implications for the resource-based view of the firm. *Research Policy*, *27*(5), 507–523.
Mueller, J. P., & Massaron, L. (2015). *Python for data science for dummies*. John Wiley & Sons.
Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A., & Sadhukhan, S. K. (2013). Cyber-risk decision models: To insure IT or not? *Decision Support Systems*, *56*, 11–26. https://doi.org/10.1016/j.dss.2013.04.004

Müller, O., Junglas, I., vom Brocke, J., & Debortoli, S. (2016). Utilizing big data analytics for information systems research: challenges, promises and guidelines. *European Journal of Information Systems*, *25*(4), 289–302.

Nahapiet, J., & Ghoshal, S. (1998). Social Capital, Intellectual Capital, and the Organizational Advantage. *The Academy of Management Review*, *23*(2), 242. https://doi.org/10.2307/259373

Nelson, R. R., Todd, P. A., & Wixom, B. H. (2005). Antecedents of information and system quality: an empirical examination within the context of data warehousing. *Journal of Management Information Systems*, *21*(4), 199–235.

Nerur, S. P., Rasheed, A. A., & Natarajan, V. (2008). The intellectual structure of the strategic management field: an author co-citation analysis. *Strategic Management Journal*, *29*(3), 319–336. https://doi.org/10.1002/smj.659

Nerur, S., Rasheed, A. A., & Pandey, A. (2015). Citation footprints on the sands of time: An analysis of idea migrations in strategic management. *Strategic Management Journal*.

Neumayer, S., & Modiano, E. (2010). Network Reliability With Geographically Correlated Failures (pp. 1–9). IEEE. https://doi.org/10.1109/INFCOM.2010.5461984

Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification.' *The Journal of Strategic Information Systems*, *24*(1), 3–14. https://doi.org/10.1016/j.jsis.2015.02.001

Nurse, J. R. C., Buckley, O., Legg, P. A., Goldsmith, M., Creese, S., Wright, G. R. T., & Whitty, M. (2014). Understanding Insider Threat: A Framework for Characterising Attacks (pp. 214–228). IEEE. https://doi.org/10.1109/SPW.2014.38

Öğüt, H., Raghunathan, S., & Menon, N. (2011). Cyber Security Risk Management: Public Policy Implications of Correlated Risk, Imperfect Ability to Prove Loss, and Observability of Self-Protection: Cyber Security Risk Management. *Risk Analysis*, *31*(3), 497–512. https://doi.org/10.1111/j.1539-6924.2010.01478.x

Olijnyk, N. V. (2015). A quantitative examination of the intellectual profile and evolution of information security from 1965 to 2015. *Scientometrics*, *105*(2), 883–904. https://doi.org/10.1007/s11192-015-1708-1

Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, *28*(6), 441–453. https://doi.org/10.1177/016555150202800601

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (Vol. 10).

Pant, G., & Sheng, O. R. L. (2015). Web Footprints of Firms: Using Online Isomorphism for Competitor Identification. *Information Systems Research*, *26*(1), 188–209. https://doi.org/10.1287/isre.2014.0563

Pant, G., & Srinivasan, P. (2010). Predicting Web Page Status. *Information Systems Research*, *21*(2), 345–364. https://doi.org/10.1287/isre.1080.0231

Pant, G., & Srinivasan, P. (2013). Status Locality on the Web: Implications for Building Focused Collections. *Information Systems Research*, *24*(3), 802–821. https://doi.org/10.1287/isre.1120.0457

Pardee, R. L. (1990). Motivation Theories of Maslow, Herzberg, McGregor & McClelland. A Literature Review of Selected Theories Dealing with Job Satisfaction and Motivation.

Park, S.-H., Huh, S.-Y., Oh, W., & Han, S. P. (2012). A social network-based inference model for validating customer profile data. *MIS Quarterly*, *36*(4), 1217–1237.

Pennebaker, J. W., & Graybeal, A. (2001). Patterns of Natural Language Use: Disclosure, Personality, and Social Integration, 5.

Pestian, J. P., Matykiewicz, P., Linn-Gust, M., South, B., Uzuner, O., Wiebe, J., … Brew, C. (2012). Sentiment Analysis of Suicide Notes: A Shared Task. *Biomedical Informatics Insights*, *5s1*, BII.S9042. https://doi.org/10.4137/BII.S9042

Peter J. Lane, & Michael Lubatkin. (1998). Relative Absorptive Capacity and Interorganizational Learning. *Strategic Management Journal*, *19*(5), 461–477.

Petrini, M., & Pozzebon, M. (2009). Managing sustainability with the support of business intelligence: Integrating socio-environmental indicators and organisational context. *The Journal of Strategic Information Systems*, *18*(4), 178–191. https://doi.org/10.1016/j.jsis.2009.06.001

Pilkington, A., & Meredith, J. (2009). The evolution of the intellectual structure of operations management—1980–2006: A citation/co-citation analysis. *Journal of Operations Management*, *27*(3), 185–202. https://doi.org/10.1016/j.jom.2008.08.001

Pinczowski, D., Ekbom, A., Baron, J., Yuen, J., & Adami, H.-O. (1994). Risk factors for colorectal cancer in patients with ulcerative colitis: a case-control study. *Gastroenterology*, *107*(1), 117–120.

Popovič, A., Hackney, R., Coelho, P. S., & Jaklič, J. (2014). How information-sharing values influence the use of information systems: An investigation in the business intelligence systems context. *The Journal of Strategic Information Systems*, *23*(4), 270–283. https://doi.org/10.1016/j.jsis.2014.08.003

Portes, A. (2014). Downsides of social capital. *Proceedings of the National Academy of Sciences*, *111*(52), 18407–18408.

Posey Garrison, C., & Ncube, M. (2011). A longitudinal analysis of data breaches. *Information Management & Computer Security*, *19*(4), 216–230.

Prell, C. (2012). *Social network analysis: History, theory and methodology*. Sage.

Provost, F., Martens, D., & Murray, A. (2015). Finding Similar Mobile Consumers with a Privacy-Friendly Geosocial Design. *Information Systems Research*, *26*(2), 243–265. https://doi.org/10.1287/isre.2015.0576

Rachel Griffith, Stephen Redding, & John Van Reenen. (2003). R&D and Absorptive Capacity: Theory and Empirical Evidence. *The Scandinavian Journal of Economics*, *105*(1), 99–118.

Raghunathan, S., & Sarkar, S. (2016). Competitive Bundling in Information Markets: A Seller-Side Analysis. *Mis Quarterly*, *40*(1), 111–131.

Ramos-Rodríguez, A.-R., & Ruíz-Navarro, J. (2004). Changes in the intellectual structure of strategic management research: a bibliometric study of theStrategic Management Journal, 1980–2000. *Strategic Management Journal*, *25*(10), 981–1004. https://doi.org/10.1002/smj.397

Ranganathan, C., & Ganapathy, S. (2002). Key dimensions of business-to-consumer web sites. *Information & Management*, *39*(6), 457–465. https://doi.org/10.1016/S0378-7206(01)00112-4

Roberts, N., Galluch, P. S., Dinger, M., & Grover, V. (2012). Absorptive capacity and information systems research: Review, synthesis, and directions for future research. *MIS Quarterly*, 625–648.

Robins, J., & Wiersema, M. F. (1995). A resource-based approach to the multibusiness firm: Empirical analysis of portfolio interrelationships and corporate financial performance. *Strategic Management Journal*, *16*(4), 277–299.

Rogers, M. (2004). Networks, firm size and innovation. *Small Business Economics*, *22*(2), 141–153.

Ronald V. Clarke. (1983). Situational Crime Prevention: Its Theoretical Basis and Practical Scope. *Crime and Justice*, 225–256.

Rook, K. S. (1984). The negative side of social interaction: impact on psychological well-being. *Journal of Personality and Social Psychology*, *46*(5), 1097.

Rouibah, K., & Ould-Ali, S. (2002). PUZZLE: a concept and prototype for linking business intelligence to business strategy. *The Journal of Strategic Information Systems*, *11*(2), 133–152.

Roussinov, D., & Chau, M. (2008). Combining information seeking services into a meta supply chain of facts. *Journal of the Association for Information Systems*, *9*(3/4), 175.

Rudolfer, S. M., Paliouras, G., & Peers, I. S. (1999). A comparison of logistic regression to decision tree induction in the diagnosis of carpal tunnel syndrome. *Computers and Biomedical Research*, *32*(5), 391–414.

Ruef, M. (2002). Strong ties, weak ties and islands: structural and cultural predictors of organizational innovation. *Industrial and Corporate Change*, *11*(3), 427–449.

Scott, J. (2012). *Social network analysis*. Sage.

Sen, R., & Borle, S. (2015). Estimating the Contextual Risk of Data Breach: An Empirical Approach. *Journal of Management Information Systems*, *32*(2), 314–341. https://doi.org/10.1080/07421222.2015.1063315

Sharma, R., Mithas, S., & Kankanhalli, A. (2014). Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems*, *23*(4), 433–441.

Shefer, D., & Frenkel, A. (2005). R&D, firm size and innovation: an empirical analysis. *Technovation*, *25*(1), 25–32. https://doi.org/10.1016/S0166-4972(03)00152-4

Shi, Z., Lee, G. M., & Whinston, A. B. (2015). Toward a better measure of business proximity: Topic modeling for industry intelligence.

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2009). Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science and Technology*, *60*(3), 571–580. https://doi.org/10.1002/asi.20994

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *Mis Quarterly*, 553–572.

Shoemaker, D. J. (2018). *Theories of delinquency: An examination of explanations of delinquent behavior*. Oxford University Press.

Shollo, A., & Galliers, R. D. (2016). Towards an understanding of the role of business intelligence systems in organisational knowing: The role of BI systems in organisational knowing. *Information Systems Journal*, *26*(4), 339–367. https://doi.org/10.1111/isj.12071

Silberman, M. (1976). Toward a Theory of Criminal Deterrence. *American Sociological Review*, *41*(3), 442. https://doi.org/10.2307/2094253

Simpson, W. (2001). QAP: The quadratic assignment procedure. In *North American Stata Users' Group Meeting, March* (pp. 12–13).

Sirgy, M. J., Efraty, D., Siegel, P., & Lee, D.-J. (2001). A new measure of quality of work life (QWL) based on need satisfaction and spillover theories. *Social Indicators Research*, *55*(3), 241–302.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*(4), 813.

Smith, P. C., & others. (1969). The measurement of satisfaction in work and retirement: A strategy for the study of attitudes.

Straub, D. W., & Welke, R. J. (1998). Coping with Systems Risk: Security Planning Models for Management Decision Making. *MIS Quarterly*, *22*(4), 441. https://doi.org/10.2307/249551

Stuart, T. E. (1998). Network Positions and Propensities to Collaborate: An Investigation of Strategic Alliance Formation in a High-Technology Industry. *Administrative Science Quarterly*, *43*(3), 668. https://doi.org/10.2307/2393679

Sundararajan, A., Provost, F., Oestreicher-Singer, G., & Aral, S. (2013). **Research Commentary** — Information in Digital, Economic, and Social Networks. *Information Systems Research*, *24*(4), 883–905. https://doi.org/10.1287/isre.1120.0472

Susarla, A., Barua, A., & Whinston, A. B. (2010). Multitask Agency, Modular Architecture, and Task Disaggregation in SaaS. *Journal of Management Information Systems*, *26*(4), 87–118. https://doi.org/10.2753/MIS0742-1222260404

Tan, A.-H., & others. (1999). Text mining: The state of the art and the challenges. In *Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases* (Vol. 8, pp. 65–70). Retrieved from http://www.ntu.edu.sg/home/asahtan/papers/tm_pakdd99.pdf

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Taylor, M. J., McNicholas, C., Nicolay, C., Darzi, A., Bell, D., & Reed, J. E. (2014). Systematic review of the application of the plan–do–study–act method to improve quality in healthcare. *BMJ Qual Saf*, *23*(4), 290–298.

Teng, J. T. C., Grover, V., & Guttler, W. (2002). Information technology innovations: general diffusion patterns and its relationships to innovation characteristics. *IEEE Transactions on Engineering Management*, *49*(1), 13–27. https://doi.org/10.1109/17.985744

Tett, R. P., & Meyer, J. P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: path analyses based on meta-analytic findings. *Personnel Psychology*, *46*(2), 259–293.

Tsai, W. (2001). KNOWLEDGE TRANSFER IN INTRAORGANIZATIONAL NETWORKS: EFFECTS OF NETWORK POSITION AND ABSORPTIVE CAPACITY ON BUSINESS UNIT INNOVATION AND PERFORMANCE. *Academy of Management Journal*, *44*(5), 996–1004. https://doi.org/10.2307/3069443

Tsai, W., & Ghoshal, S. (1998). SOCIAL CAPITAL AND VALUE CREATION: THE ROLE OF INTRAFIRM NETWORKS. *Academy of Management Journal*, *41*(4), 464–476. https://doi.org/10.2307/257085

Tun Thura Thet, Na, J.-C., & Khoo, C. S. G. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of Information Science*, *36*(6), 823–848. https://doi.org/10.1177/0165551510388123

Vance, A., Siponen, M., & Pahnila, S. (2012). Motivating IS security compliance: Insights from Habit and Protection Motivation Theory. *Information & Management*, *49*(3–4), 190–198. https://doi.org/10.1016/j.im.2012.04.002

von Solms, R., & van Niekerk, J. (2013). From information security to cyber security. *Computers & Security*, *38*, 97–102. https://doi.org/10.1016/j.cose.2013.04.004

Vroom, C., & von Solms, R. (2004). Towards information security behavioural compliance. *Computers & Security*, *23*(3), 191–198. https://doi.org/10.1016/j.cose.2004.01.012

Wakefield, R. (2013). The influence of user affect in online information disclosure. *The Journal of Strategic Information Systems*, *22*(2), 157–174. https://doi.org/10.1016/j.jsis.2013.01.003

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations* (pp. 115–120). Association for Computational Linguistics.

Wang, L., & Zajac, E. J. (2007). Alliance or acquisition? a dyadic perspective on interfirm resource combinations. *Strategic Management Journal*, *28*(13), 1291–1317. https://doi.org/10.1002/smj.638

Wang, T., Kannan, K. N., & Ulmer, J. R. (2013). The Association Between the Disclosure and the Realization of Information Security Risk Factors. *Information Systems Research*, *24*(2), 201–218. https://doi.org/10.1287/isre.1120.0437

Wang, X. (Shane), Bendle, N. T., Mai, F., & Cotte, J. (2015). The *Journal of Consumer Research* at 40: A Historical Analysis. *Journal of Consumer Research*, *42*(1), 5–18. https://doi.org/10.1093/jcr/ucv009

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.

White, H. D. (1990). Author co-citation analysis: Overview and defense. *Scholarly Communication and Bibliometrics*, *84*, 106.

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the Association for Information Science and Technology*, *32*(3), 163–171.

White, H. D., & Griffith, B. C. (1982). AUTHORS AS MARKERS OF INTELLECTUAL SPACE: CO-CITATION IN STUDIES OF SCIENCE, TECHNOLOGY AND SOCIETY. *Journal of Documentation*, *38*(4), 255–272. https://doi.org/10.1108/eb026731

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science*, *49*(4), 327–355.

Wiatrowski, M. D., Griswold, D. B., & Roberts, M. K. (1981). Social Control Theory and Delinquency. *American Sociological Review*, *46*(5), 525. https://doi.org/10.2307/2094936

Wiemer-Hastings, P., Wiemer-Hastings, K., & Graesser, A. (2004). Latent semantic analysis. In *Proceedings of the 16th international joint conference on Artificial intelligence* (pp. 1–14). Citeseer.

Wignaraja, G. (2002). Firm size, technological capabilities and market-oriented policies in Mauritius. *Oxford Development Studies*, *30*(1), 87–104.

Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management*, *17*(3), 601–617.

Xu, K., Liao, S. S., Li, J., & Song, Y. (2011). Mining comparative opinions from customer reviews for Competitive Intelligence. *Decision Support Systems*, *50*(4), 743–754. https://doi.org/10.1016/j.dss.2010.08.021

Yan, E., Ding, Y., Milojević, S., & Sugimoto, C. R. (2012). Topics in dynamic research communities: An exploratory study for the field of information retrieval. *Journal of Informetrics*, *6*(1), 140–153.

Yin, D., Bond, S., & Zhang, H. (2013). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews.

Zahra, S. A., & George, G. (2002). Absorptive Capacity: A Review, Reconceptualization, and Extension. *The Academy of Management Review*, *27*(2), 185. https://doi.org/10.2307/4134351

Zhao, D., & Strotmann, A. (2008). Evolution of research activities and intellectual influences in information science 1996-2005: Introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology*, *59*(13), 2070–2086. https://doi.org/10.1002/asi.20910

Zhao, D., & Strotmann, A. (2015). *Analysis and visualization of citation networks*. San Rafael: Morgan & Claypool.

Zhao, J., Gou, L., Wang, F., & Zhou, M. (2014). PEARL: An interactive visual analytic tool for understanding personal emotion style derived from social media (pp. 203–212). IEEE. https://doi.org/10.1109/VAST.2014.7042496

Zheng, Z. (Eric), Fader, P., & Padmanabhan, B. (2012). From Business Intelligence to Competitive Intelligence: Inferring Competitive Measures Using Augmented Site-Centric Data. *Information Systems Research*, *23*(3-part-1), 698–720. https://doi.org/10.1287/isre.1110.0385

Zhu, D., Prietula, M. J., & Hsu, W. L. (1997). When processes learn: Steps toward crafting an intelligent organization. *Information Systems Research*, *8*(3), 302–317.