THE EFFECTS OF NETWORK STRUCTURE ON SUPPLY CHAIN RISK PROPAGATION: A SIMULATION STUDY

BY MYLES D. GARVEY

A dissertation submitted to the

Graduate School—Newark

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Management

Written under the direction of

Sengun Yeniyurt

and approved by

_____

_____

_____

_____

Newark, New Jersey

October, 2018

© 2018

Myles D. Garvey

ALL RIGHTS RESERVED

ABSTRACT OF THE DISSERTATION

The Effects of Network Structure on Supply Chain Risk Propagation: A Simulation Study

by Myles D. Garvey

Dissertation Director: Sengun Yeniyurt

Disruptions frequently occur in supply chains and hence pose problems for practitioners of whom seek to design efficient and resilient supply networks. Events that lead to such disruptions are commonly referred to as Supply Chain Risks. While many have studied the effect that risk has on consequences of certain decisions made, few have attempted to explore the effects that the inherent structure of a supply network has on the global and local risk of a supply chain. In addition, few have studied the effects that global and local network structure have on risk propagation to individual firms and connections within the network. In this dissertation, I explore the effects that inherent global and local network structural characteristics, operationalized via graph-theoretic measures as well as local centrality measures, have on global and local risk propagation. I do so by employing a simulation study to generate a very large data set of networks and various risk distributions. The primary risk measurement tool used in this study are Bayesian Networks. Last, I explore the effect that various strategic decisions have on global and local risk, and how such decisions can alter the current state of risk in supply networks.

"All we can know is that we know nothing. And thats the height of human wisdom"

-Leo Tolstoy, War and Peace, Pierre

Acknowledgements

This dissertation would not be possible without the help, wisdom, motivation and compassion from many individuals, both divine and concrete, that are in my life. I would like to take careful time to acknowledge each one personally.

First and foremost, I would like to thank the Lord Jesus Christ. Many know me as not much of a religious man, but the truth is, I am deeply faithful within and privately. The struggle of which myself and my family have endured throughout the prior few years would not have been possible to survive without the path that my Lord and Savior has set forth in front of me. Where many days and nights I felt lost, the Lord has always guided me back on the path to righteousness and hope. Without such divine intervention during my deep pitfalls in life, I do not know how else I would have survived.

I would like to acknowledge my entire dissertation committee. Without their guidance and wisdom, this dissertation would not at all be possible. They have ensured throughout this entire process that I stay on target with my goals and objectives, and have challenged me intellectually and philosophically. Without them, I certainly would not have been able to ascertain the level of knowledge I have to this day. Their suggestions and willingness to help has proven invaluable.

I also would like to acknowledge Steven Carnovale, my dear friend of now 16 years. Without his support in nearly every dimension of my life, I honestly do not know how I would stand here today. He is a true friend, one that any man can hope to have, and I consider him another brother of mine. His infinite wisdom and persistent challenging of my own ideas and beliefs has allowed myself to further grow intellectually, which I would find quite near impossible to do otherwise. The magnitude of his compassion for myself and my family is ineffable. I am eternally indebted to his compassion and

friendship.

In addition, I would like to acknowledge Steven's wife, Jessica Carnovale. She has been like a sister to me, always making sure, despite my own stubbornness, that I seek the proper help when needed. She has always been there for me, and I see her no different than I would a biological sister. Many nights over coffee and many days of sharing meals with both her and her husband was certainly a cornerstone of my life that is quite unforgettable. And so, I would like to acknowledge her for ensuring that I stay on track with my goals, because like her husband Steven, she too wanted to ensure that I always sought out what was best and healthy for myself, just as any sister would.

I would also like to acknowledge Sengun Yeniyurt and Michael Katehakis. Both have served as advisors of mine throughout the years, and have guided me on both a personal and professional level. Their consistent reminders and long discussions on a variety of topics, both personal and professional, has given myself the motivation to press further in this long and arduous path. Their wisdom and guidance is truly rare in this world, and I am honored to have been openly embraced as a student of theirs'.

In addition, I would like to acknowledge Dr. Mark Turpin, my previous mentor at the University of Hartford in the Mathematics Department. Private discussions with him throughout the years have only further provided additional motivation for myself to enter the path into academia. Countless hours of discussion in his office, on a variety of topics in Mathematics, all those years ago, had laid and cemented the mechanical foundation of my own mathematical insight. Without him, and the cementing of that foundation all those years ago, I truly would have found it difficult to justify this dissertation on a mechanical level.

I would like to acknowledge my future wife, Shannon Kelly. Shannon has been nothing short of patient and supportive during this entire process. Without her unwaivering love and compassion towards myself and my own career, I honestly feel the path to this day would have been greatly extended. She truly is my one and only love. It has been an arduous journey for both of us these past few years. We can eventually look back on this text, forever etched in the annals of time, when we are both old and gray, and recall all the tough-times we have endured together, and remember that no matter the

magnitude of the challenge, we will always make it through on the other side, together.

I would also like to acknowledge Shannon's family. Her father, Patrick Kelly, has been like a second father to me. His wisdom and life advice as well as emotional support throughout this entire process is greatly appreciated. It's an honor to have been accepted into his family. The rest of Shannon's family has also provided a magnitude of compassion and support to myself that certainly deserves acknowledgement through this process. Without them, my time and patience for working on my degree would certainly have been short-lived.

My brother Douglas Garvey, and one of my best friends of 28 years, Raymond Yeszefski, both deserve proper acknowledgement as well. Without their continued persistence for me to dip my head out of the books and take a break, I most likely would have burnt out. Their support and friendship mean the world to me. Their passion for music, film and the arts gave myself an intellectual break, and allowed myself to rest my mind for this very important endeavor. A man cannot ask for a better brother and a friend.

Last, and most importantly, I would like to thank my parents, Marianne and Myles R. Garvey. My parents have guided me throughout the years. I had learned my own work-ethic from both of them, observing as a child their own life struggles. Without their advice, wisdom and guidance, I would most certainly not be the individual I am today. Since a young child I was lectured that hard-work, rather than time, will always pay-off, and the only way that one fails is if one gives up. That lesson was personified through their own struggles of which I had directly observed throughout my own life. Without their direct and indirect lessons, I don't know if I would have had the motivation to continue down this long and arduous path. Their love and compassion for me is second to no one. They will always remain an inspiration for me, and I can only hope to continue to aspire to be just like them.

Dedication

This dissertation is dedicated to my children, Cadence Harmony Kelly Durning and Myles Patrick Garvey. I love both of you with all of my heart. Without your funny faces, crazy antics, and infinite, NEVER-ENDING questions, I don't know if I really could have completed this work.

I want you both to know that no matter the challenge life throws your way, with enough motivation, support, and hard-work, anyone can accomplish their goals. Also, knowledge is something that shouldn't be a trivial fleeting moment in time, but rather a deep, inquisitive, and exciting journey, not through mindless regurgitation of fact that so many of whom lay claim to knowledge do, but rather through logic, with equal levels of careful consideration and skepticism of factuals and counter-factuals. Don't EVER stop asking questions or challenging others. Yes, do so with respect, but always ensure you challenge your educators, for that is how real knowledge is attained.

So I end this dedication on this final bit of advice my children. Do not let negativity from others dissuade you. While I had much support from others, I equally had negativity from some. I want you to know that this dissertation is the manifestation of one of my many dreams in life. It is proof that no matter how trifling your dreams may appear to others, that you should still pursue them, no matter how close to the heavens and stars they may appear to you or others. Last, do not let failure get in the way of your success. Embrace failure, it is the most important lesson a man or woman can learn in life. I can attest, for this dissertation had hundreds of failures before it's thankful and eventual success. If there is any proof you need of this fundamental principle in life, look no further than this dissertation.

With love for both of you,

Dad

Table of Contents

Abstract ii

Acknowledgements iv

Dedication vii

List of Tables xiii

List of Figures xv

1. Introduction 1

1.1. Background and Motivation for the Dissertation 2

1.1.1. Supply Chains 4

1.1.2. Supply Chain Disruption and Risk 7

1.1.3. Structural Network Characteristics and Supply Chain Risk . . . 11

1.2. Research Objectives and Inquiries 12

1.2.1. Conceptual Motivation 13

1.2.2. Mathematical Motivation for the Dissertation 15

2. Literature Review 26

2.1. Supply Chain Risk and Disruption 26

2.1.1. Supply Chain Disruptions 26

2.1.2. Supply Chain Risk and Risk Management 30

The Need for Supply Chain Risk Management (SCRM) 30

SCRM Frameworks: Evolution and Current State 33

Risk Taxonomies: A Lack of Consensus 36

Supply Chain Risk Identification 39

Supply Chain Risk Assessment 41

Some Common Mitigation Strategies 42
2.1.3. Supply Chain Risk Measurement 44
Prior Attempts to Quantify Supply Chain Risk 45
2.1.4. Supply Chain Risk Propagation (SCRP) 51
Risk Propagation vs. Disruption Propagation 51
A Review of Risk Propagation 53
2.2. Fundamentals of Bayesian Networks 56
2.2.1. Conditional Probability and Independence 56
2.2.2. Bayesian Networks: A Model of Belief Propagation 57
2.2.3. Learning Bayesian Networks 59
2.2.4. Belief Inference and Propagation 60
2.3. Supply Network Structure 65
2.3.1. The Basic Supply Network Model 66
2.3.2. The Risk Graph 67
2.3.3. Measuring Network Structure 71
Context-Dependent Conceptualization of Structure 73
Partial-Context-Free and Context-Dependent Classification and Taxonomy 76
Graph-Theoretic Based Characterization 81
2.3.4. Network Structure Dynamics in Supply Chains 86
2.3.5. Why Bother to Classify? 87
2.4. Current Research Gaps in the Literature 88
2.4.1. Context-Free Network Structure Conceptualization 88
2.4.2. Network-Based Hierarchical Levels of Structural Analysis 91
2.4.3. Risk Propagation 91
2.4.4. Intersection of Risk Propagation and Network Structure 92
2.4.5. Supply Chain Horizon and Risk Propagation 93
2.4.6. Hierarchical Levels of Structural Analysis and Risk 93
2-Ego Network Risk 93

Ego-Network Risk 94
Node and Arc-Level Risk 94
2.4.7. Research Questions 95

3. Theory Development 96
3.1. Supply Chains and Network Structure 99
3.1.1. The Level-of-Analysis for Networks: A Contextual Perspective . 99
3.1.2. The Level-of-Analysis for Networks: A Non-Contextual Perspective 104
3.1.3. Network Characterizations: A Mess of Context 108
3.1.4. The Fundamental 5-Dimensions: Towards Characterizing Networks 111
Centrality 113
Connectedness 114
Clustering 116
Feedback/Thickness 118
Distance 120
3.2. Risk Measurement and Risk Propagation: How Risk "Flows" 121
3.2.1. A Philosophical Discussion on Risk Measurement 121
3.2.2. The Line of Demarcation between Supply Chain Risk and Supply Chain Risk Propagation 127
3.2.3. The Problem of Deterministic Propagation 128
3.2.4. The Invisible Path of Causality 129
3.2.5. Balancing Model Complexity and Error in Risk Propagation . . 132
3.2.6. Constructs of Supply Chain Risk Propagation 137
Propagation Diagnostics: Risk Significance 139
Propagation Causality: Risk Contribution 141
Propagation Change: Risk Velocity 142
3.2.7. Methods of Operationalization for Risk Propagation Constructs . 143
3.3. Hypothesis Development 145
3.3.1. Conceptual Development 145

2-Ego Network Risk 145
1-Ego Network Risk 153
Node/Arc Level Risk 156
Cross-Level Risk Effects 158

4. Research Design 162
4.1. Population and Variable Definition 162
4.1.1. Description of the Population of Interest 162
4.1.2. Dependent Variables 163
4.1.3. Independent Variables 164
Control Variables 169
4.2. Design of the Simulation 169
4.3. Description of Resulting Datasets 172
4.4. Empirical Models 173

5. Analysis and Results 175
5.1. Descriptive Statistics 175
5.2. Factor Analysis 184
5.2.1. 1-Ego-Network and 2-Ego-Network Factor Analysis 184
5.2.2. Node and Edge Factor Analysis 189
5.3. Structural Difference Testing 200
5.4. Causal Analysis 202
5.5. Results 214
5.5.1. 2-Ego-Network Risk 214
5.5.2. 1-Ego-Network Risk 216
5.5.3. Local Risk 218
5.5.4. Cross-Risk and Structural Differences 219
5.6. Summary of Hypothesis Test Results 220

6. Discussion, Limitations and Managerial Implications 222

6.1. Discussion of Results 222
6.2. Contributions of this Research 230
6.3. Limitations of Research 234
6.4. Managerial Implications 236
6.5. Future Research and Open Research Questions 237

7. Conclusion 240

Appendix A. R-Code for Simulation 243

Appendix B. R Code for Data Analysis 279

References 322

List of Tables

2.1. SCRM Framework Dimensions 36

3.1. Conditional Probability Table of Disruption at Retailer Given Manufacturer and Raw Material 135

3.2. Conditional Probability Table of Disruption at Raw Material 135

5.1. Summary Statistics of the Raw Network Data 176

5.2. Summary Statistics for The 2-Ego-Network Original Data-Set 177

5.3. Summary Statistics for The 1-Ego-Network Original Data-Set 179

5.4. Summary Statistics for The Node/Edge Original Data-Set 180

5.5. The 2-Ego-Network Correlation Matrix. Gray cells indicate $p < 0.05$. . . 181

5.6. The 1-Ego-Network Correlation Matrix. Gray cells indicate $p < 0.05$. . . 182

5.7. The Local Correlation Matrix. Gray cells indicate $p < 0.05$. 183

5.8. The 1-Ego-Network and 2-Ego-Network Factor Loadings from the EFA. 186

5.9. The EFA Loadings on the Node/Edge data set 192

5.10. The 2-Ego-Network Risk CFA Statistics 194

5.11. The 2-Ego-Network Risk CFA Loadings 195

5.12. The 1-Ego-Network Risk CFA Statistics 196

5.13. The 1-Ego-Network Risk CFA Loadings 196

5.14. The 2-Ego-Network Factored Correlation Matrix. Gray cells indicate $p < 0.05$. 197

5.15. The 1-Ego-Network Factored Correlation Matrix. Gray cells indicate $p < 0.05$. 198

5.16. The Local-Network Factored Correlation Matrix. Gray cells indicate $p < 0.05$. 199

5.17. Summary Statistics for The Factored 2-Ego-Network Data-Set. 200

5.18. Summary Statistics for The Factored 1-Ego-Network Data-Set. 202
5.19. Summary Statistics for The Factored Node/Edge Data-Set. 203
5.20. Results of One-Tailed Structural Comparison. 203
5.21. The VIFs for the variables in the 2-Ego-Network Model. 214
5.22. The VIFs for the variables in the 1-Ego-Network Model. 215
5.23. The VIFs for the variables in the Local-Network Model. 216
5.24. Results from SUR Estimation of 2-Ego-Network System 217
5.25. Results from SUR Estimation of 1-Ego-Network System 217
5.26. Results from SUR Estimation of Node/Edge System 218
5.27. The set of hypotheses and their results. 221

List of Figures

1.1. Four different directed networks, each one of which has 9 nodes. 17

1.2. The Expected Location Risk Contribution Factor (ELRCF) plotted against the 4 different network's distance and centrality measures. 20

1.3. The Coefficient of Variation of the Expected Location Risk Contribution Factor (ELRCF) plotted against the 4 different network's distance and centrality measures. 22

1.4. Node-Level Risk plotted against the node-level centrality measures. . . 23

2.1. An example of a Bayesian Network (Garvey, Carnovale, & Yeniyurt, 2015). 58

3.1. The 0-Ego-Network (in yellow area), 1-Ego-Network (in red area), 2-Ego-Network (in green area), and 3-Ego-Network (in blue area) of node n_1. 106

3.2. The 0-Ego-Network (in yellow area), 1-Ego-Network (in red area), 2-Ego-Network (in green area), 3-Ego-Network (in blue area), and the 4-Ego-Network (in gray area) of node n_7. 107

3.3. Caption place holder 117

5.1. The scree plot for the EFA on the combined 2-Ego-Network and 1-Ego-Network data. 185

5.2. The final loadings of the 2-Ego-Network and 1-Ego-Network data. . . . 187

5.3. The scree plot for the EFA on the combined Node and Edge data. . . . 190

5.4. The final factor loadings for the Node and Edge data. 191

5.5. Densities for each Construct by Level of Analysis 201

5.6. The 2-Ego-Network Densities for the dependent variables (above) and the constructs (below). 204

5.7. The 1-Ego-Network Densities for the independent variable risk factors (above) and the constructs (below). 205
5.8. The 1-Ego-Network Densities for the dependent risk variables (above) and the Local dependent risk variables (below) 206
5.9. The Local Densities for the independent risk variables (above) and the local densities for the constructs 207
5.10. 2-Ego-Network Risk (CV of Velocity) plotted against 2-Ego-Network Structure Constructs 209
5.11. 1-Ego-Network Risk (CV of Velocity) plotted against 1-Ego-Network Structure Constructs 210
5.12. 1-Ego-Network Risk (CV of Velocity) plotted against 2-Ego-Network Risk Constructs 211
5.13. Local-Network Risk (CV of Velocity) plotted against Local-Network Structure Constructs 212
5.14. Local-Network Risk (CV of Velocity) plotted against 1-Ego-Network Risk Constructs 213
6.1. 2-Ego Network Risk Structural Model 224
6.2. 1-Ego Network Structural Model 225
6.3. 1-Ego Network Risk Model 226
6.4. Local Network Structural Model 227
6.5. Local Network Risk Model 228

Chapter 1

Introduction

In April of 2016, Airbus had reported a 50% drop in their profit due to the delay in shipment of orders. Reports at the time indicated that this was due to the difficulty in both the design and execution phase of production. Further compounding the issue was increased demand alongside high recurring costs. In addition, delay in shipment of engines manufactured from Pratt and Whitney had itself led to additional delays in shipment of the fully assembled planes (Wall, 2016).

It was only a few days prior to the release of Airbus's profits that a deadly 6.5 magnitude earthquake had struck Japan, leaving nine dead as a result. A spokesman from Sony stated that they were assessing any potential damage at a nearby image-sensor plant in Kumamoto, whose products are used in the manufacture of Apple's iPhone (Landers & Karioka, 2016) Not only was Sony affected by the earthquake, but so too was Toyota. It was only a few days after reports of the quake that Toyota announced it was shutting down 26 car assembly lines due to a disturbance in production from a supplier. Reports indicated that Honda and Nissan also fell victim to the the disturbance in production. However, the shutdown in Toyota's assembly lines was nationwide, unlike it's competitors, due to it's lean supply chain. Reports had indicated that the shutdown was likely to cost Toyota tens of billions of yen (Kubota, 2016).

It wasn't only Japanese car manufacturers that had to respond with action to the quake, however. Days after the quake, GM announced it was shutting down production in assembly plants in Tennessee, Canada, Ohio and Kansas. Their intention was to thwart the disruption that had occurred in Japan just a few days prior. Unlike Toyota, however, reports indicated that GM was not anticipating any performance or financial loss (Nagesh, 2016).

Nearly a month before the earthquake, on St. Patrick's Day, it was reported that Aeropostle was seeking to explore a resolution to a dispute it had with one of it's primary suppliers, MGF Sourcing. The conflict began with allegations from Aeropostle that MGF had not fulfilled a 10-year sourcing agreement. The alleged violation of the agreement, according to Aeropostle, led to a lack of availability of certain merchandise, and had warned would lead to a liquidity constraint. Their stock fell by nearly 47% after reporting a loss for it's 13th quarter of continual loss (Safdar & Minaya, 2016).

Most recently, at the time of this writing, many automotive companies feared the impact of the tariffs placed on China by President Trump, and China's retaliatory tariffs placed on America, if a trade agreement between the two countries was not settled. Different automotive companies responded differently near the enactment of the tariffs. Ford stated that it would cut into it's margins while keeping it's prices the same. Lincoln allegedly had the same response. Unlike Ford and Lincoln, Tesla intended to raise its prices. Many other companies in both America and China had feared the potential consequences of the tariffs (Kubota, Deng, & Craymer, 2018).

The prior anecdotes recounted here are only a handful of many supply chain disruptions that occur in firms across the globe every year. They illustrate that disruptions are not persnickety about industry, firm, country or type. They are bound to occur at any moment in time due to an unforeseen event or sequence of events. Supply chain disruptions can occur due to labor union disputes, contractual disputes between buyers and suppliers, natural disasters, and geopolitical events, among many other reasons. The question for supply chain practitioners then is, how shall we manage them?

1.1 Background and Motivation for the Dissertation

As the aforementioned anecdotes illustrate, supply chain managers design their networks not only for competitive advantage or performance, but also for handling disruptions. This of course depends on their strategic objectives of having their supply chains as being more responsive or efficient. In the case of Toyota, their assembly lines were nationwide, and as a result, a shutdown of all them was necessary not only due

to the earthquake itself, or the subsequent disruption, but also due to it's structural design. The anecdote illustrates that despite Toyota being required to thwart it's nationwide assembly lines, Nissan and Honda had only suffered a shutdown for a short period of time and in only a constrained number of it's lines, with minimal loss. This is further personified in the instance of GM, whose shutdown occurred in four different locations due to an incident that happened thousands of miles away. Yet, their loss was considered trifling.

The earthquake story partly supports the notion that structural characteristics of a firm's supply network and strategic decision making in establishing and maintaining buyer/supplier relations plays a role in determining the potential loss to a firm in the event of a disruption. Logic would indicate that if it did not, the earthquake inherently would have ceased production of all manufacturing facilities, and would have resulted in a loss of similar magnitude across all the automotive firms. This was not the case, however, in the recounted anecdote. An interesting question that arises out of this anecdote, as well as the others, is: "how should managers account for supply chain structure in their strategic decision making?"

The managerial implication of answering this question is important for practitioners that seek to prepare risk-mitigation strategies in their supply networks, while engaging in other strategic activities. Indeed, it should be understood that supply chain risk management intersects strategic decision making. For if it is not, a firm is being myopic to the possible detrimental effects in the event of a disruption upstream, and a previously missed trade-off may come to fruition. While costs and performance may be lower and higher, respectively, for a potential supplier, their risk of disruption may be considerably high due to connections they themselves have to other high risk-of-disruption firms.

Hence, this dissertation seeks to further explore this inquiry for the primary purpose of expanding strategic decision making. In addition, knowing how supply network structure plays a role in determining a firm's risk of disruption extends to other activities of the firm that are not necessarily strategic in nature. Production, inventory, financial, logistic and information processes all partly depend on the actors within the supply network itself. Knowledge of the nature of the relationship between the constructs of

structure and risk would advance the body of knowledge of supply chain management across multiple dimensions and units of analysis of the firm.

In the remainder of this section, I will offer further motivation for the primary topic of this dissertation. I will first do so by motivating the definition of a supply chain. Next, I will define and discuss the basic problem of supply chain disruptions and risk of disruption. Last, I will discuss the motivations behind supply chain risk and its potential connection with network characteristics.

1.1.1 Supply Chains

I begin the motivation for the primary research question by carefully defining some of the concepts and terms of which is commonplace in the extant literature. The term *supply chain* is actually quite misleading, as it would indicate a simple chain structure. Early researchers in the field viewed supply chain constructs as an extension of logistics activities, dyadic relationship management based on resource dependency theory, and as an extension of the value chain (Stevens, 1989; Oliver, Webber, et al., 1982; Houlihan, 1985). Indeed, Stevens defined the supply chain as "the connected series of activities which is concerned with planning, coordinating and controlling material, parts and finished goods from suppliers to the customer" (Stevens, 1989, pg. 3). Unlike the researchers before him (Oliver et al., 1982; Houlihan, 1985), however, he defined the supply chain as comprising of two types of flows, namely "material and information"(Stevens, 1989, pg. 3).

Soon thereafter, however, researchers began to understand that supply chains are actually much more complex than a simple chain (Ellram, 1991; Cooper, Lambert, & Pagh, 1997; Lambert & Cooper, 2000; Cooper & Ellram, 1993). Indeed, early researchers recognized that the supply chain is "not just a chain of businesses with one-to-one, business-to-business relationships, but a network of multiple businesses and relationships" (Lambert, Cooper, & Pagh, 1998, pg. 1). As such, the tools of mathematical graph theory were seen as applicable in order to gain an understanding of the complexity of supply networks (Choi & Hong, 2002; Borgatti & Everett, 2006; Choi & Krause, 2006). This was later extended to include the use of social network theory

to study the vast and complex inter-dependencies between actors in the supply chain (Borgatti & Li, 2009; Autry & Griffis, 2008; Carter, Leuschner, & Rogers, 2007; Carter, Ellram, & Tate, 2007).

It wasn't very long for research in the supply chain and supply chain management literature to take off. Various areas of research were being explored such as buyer-supplier relationship management, supply chain integration, supply chain sustainability, green supply chains, supply chain risk management, supply chain process frameworks, supply chain complexity, supply chain robustness and resiliency, procurement, and supplier selection (Croom, Romano, & Giannakis, 2000; Burgess, Singh, & Koroglu, 2006; D. C. Ho, Au, & Newton, 2002). Many of these areas of research were empirical validations of prior proposed theory (Croom et al., 2000). However, the definition of a supply chain still varied throughout the literature, as well as that of the management of supply chains, along with the various constructs within.

Researchers began to take note that while various branches of the supply chain literature were being explored, a solid theoretical foundation seemed to still be lacking (I. J. Chen & Paulraj, 2004). Various papers rested their theoretical arguments on a variety of ad-hoc, but still mainly related theoretical lens, such as the resource-based view (RBV)(Rungtusanatham, Salvador, Forza, & Choi, 2003; Barney, Wright, & Ketchen Jr, 2001; Miles & Snow, 2007), the knowledge-based view (KBV)(Miles & Snow, 2007; Hult, Ketchen Jr, & Slater, 2004), transaction cost economics (TCE) (Ireland & Webb, 2007), resource dependency theory (RDT)(Ketchen Jr & Hult, 2007; Ireland & Webb, 2007), the practice-based view (PBV) (Bromiley & Rau, 2014; Carter, Kosmol, & Kaufmann, 2017) and resource orchestration theory (ROT) (Dhanaraj & Parkhe, 2006; Hinterhuber, 2002). While these are still useful for modern research, some in the extant literature felt that a collection of theoretical and axiomatic constructs, specifically designed for supply chain theory, should be formulated to serve as a foundation for a more supply-chain specific theoretical lens (I. J. Chen & Paulraj, 2004; Carter, Rogers, & Choi, 2015; Carter et al., 2017). The proposed framework by (Carter et al., 2015) provided just that, along with a well-defined and grounded set of constructs that researchers can use to provide for consistency across the literature

(Carter et al., 2015).

The (Carter et al., 2015) framework is not intended to replace previous theoretical lens such as the RBV, rather it serves as a compliment to existing frameworks when the research is supply chain centric. The primary issue they found with the modern stream of supply chain research is that "we have failed to develop a theory of what we are managing–a theory of the supply chain"(Carter et al., 2015, pg. 89). They argue that supply chain construct definition needs to be placed on a solid theoretical foundation, and they do so in their research. I shall adopt the same definitions for this dissertation that they have proposed.

First, (Carter et al., 2015) defines a supply chain as "a network, consisting of nodes and links" (Carter et al., 2015, pg. 90). This definition is in line with the aforementioned literature as well as the modern literature in supply chains. Hence, I shall use the term "supply chain" and "supply network" synonymously. Second, (Carter et al., 2015) continues to argue that "the supply chain as a network operates as a complex adaptive system, where every agent grapples with the tension between control and emergence."(Carter et al., 2015, pg. 91) They argue that much of the supply chain literature deals with agents, where the nodes in the aforementioned definition represents a generic agent, of whom looks upstream and downstream, with a constrained view of clarity, in order to make decisions to control resources or accountability. The manifestation of these decisions made due to the lack of clarity therefore is a complex adaptive system, where the network is dynamic and ever-changing. Their definition of "agent" leaves open the ability to define a wide range of different unit of analyses for researchers, be it at the firm-level, the location-level, or other levels of analysis.

They continue their framework by contending that supply chains are defined in a relative manner, particularly to both a specific product and a specific focal firm. Again, this definition of "the supply chain" leaves enough room open for different levels of analysis. In addition, they argue there are always two types of chains that are embedded in the existence of any supply network that is defined: a physical chain and a support chain. The physical chain is much like the traditional supply chain: a network of nodes that add value to the product or service. The support chain is a collection of

nodes that do not add value but are necessary for the physical chain to operate. They use financial and information flows as an example. While finance and information do not necessarily add value to the physical product or service, they are necessary for the proper movement of the product through the physical chain. Last, (Carter et al., 2015) argues that the supply chain itself is "bounded by the visible horizon of the focal agent" (Carter et al., 2015, pg. 93). They submit that supply chains are viewed from their focal agents, and as the distance from the focal agent increases, the horizon for said agent becomes attenuated. Thus, the boundaries of the supply network, they argue, are dependent on the "awareness" of the focal agent.

In summary, I adopt (Carter et al., 2015)'s definition of *the supply chain* as being a network of nodes and arcs that together form a complex adaptive system comprising of a variety of agents, relative to a well-defined focal agent and product, of which has two embedded networks, each respectively handling the product flow, and the support for the product flow, through other agents, where the horizon is bounded by the focal agent's self-awareness.

1.1.2 Supply Chain Disruption and Risk

The aforementioned anecdotes illustrate that supply chains are susceptible to "breaking down". That is, the various processes within a supply chain change to an unexpected state that is not common. This could include a shortage of inventory due to higher demand, a late shipment from a supplier, a production process whose flow rate or time is suddenly altered, or anything else that thwarts the movement of the product or service through the various actors within the supply network.

Supply chain disruption and risk management have been a topic for researchers for two decades (Chopra & Sodhi, 2004; Denardo & Lee, 1996; G. A. Zsidisin & Ritchie, 2009; Paulsson, 2004; C. S. Tang, 2006; P. R. Kleindorfer & Saad, 2005). Just as with the definition of the supply chain, there are various definitions for "supply chain disruption" and "supply chain risk". I will conduct a more thorough literature review of this in the next chapter. For the time being, in this section, I would like to clearly define the terms "disruption" and "supply chain risk", as well as discuss the common

theme across various frameworks for their management that have been proposed.

Some authors have confounded the definitions of "risk" and "disruption" to be one in the same. For example, Hendricks states that "there seems to be widespread recognition that such disruptions have the potential to cause significant negative economic impacts" (Hendricks & Singhal, 2005b, pg. 35). That is, he is defining a disruption as a "potential", which would be interpreted as a risk. In a later study, he mentions that disruptions themselves "cause" risk.

Others, however, separate the construct of "disruption" and "risk" (Garvey, Carnovale, & Yeniyurt, 2015; DuHadway, Carnovale, & Hazen, 2017), where one is an event that *could* occur (the risk) and the other is a manifestation of that event (the disruption) (DuHadway et al., 2017). This distinction is important, as it allows for a proper classification of disruptions and the distinction between an event that is preventing the normal flow of goods, versus a potential event that could manifest into a disruption. That is, this distinction of the constructs of risk and disruption allow for a well-grounded framework to determine the antecedents of a disruption.

Supply chain disruptions are "unplanned and unanticipated events that disrupt the normal flow of goods and materials within a supply chain" (Craighead, Blackhurst, Rungtusanatham, & Handfield, 2007, pg. 132 ; Garvey et al., 2015; P. R. Kleindorfer & Saad, 2005). Equally important to defining a supply chain disruption is to define supply chain risk, which is considered to be the antecedent of a supply chain disruption (DuHadway et al., 2017). There is, however, much debate over a precise definition of "supply chain risk". Some authors characterize it in terms of consequences (Hendricks & Singhal, 2005b), where others characterize it in terms of likelihood (Trkman & McCormack, 2009; Garvey et al., 2015; P. R. Kleindorfer & Saad, 2005).

In actuality, a supply chain risk is much more multi-faceted than either case. Zsidisin, who refused to offer a single definition of supply chain risk in his research, states that they "have deliberately chosen not to promote a definition for the term risk for two reasons. Firstly, seeking agreement on a definition has proved problematic for most fields of study across significant time periods. Secondly, prescribing a particular definition is likely to prove counter-productive in generating and encouraging the different

perspectives and approaches adopted by our co-authors" (G. A. Zsidisin & Ritchie, 2009, pg. 4). He instead enumerates a list of 4 dimensions along which the construct of supply chain risk can be characterized.

Despite Zsidisin's objections being unrelated to others' definitions of risk, it is worthy to note that some have adopted a definition of supply chain risk with respect to three primary dimensions: the likelihood of the event, the severity of the event, and the causal pathways that could lead to the event (Khojasteh-Ghamari & Irohara, 2018). Indeed, "the two main elements, which are probability of occurrence and impact of risk, are critical to determining the necessity of an action to combat a risk in a supply chain" (Khojasteh-Ghamari & Irohara, 2018, pg. 8). They go on to offer a mathematical definition of supply chain risk (Khojasteh-Ghamari & Irohara, 2018; Ritchie & Brindley, 2007):

$$\text{Risk} = \text{Likelihood} \times \text{Severity} \times \text{Detection}$$

The issue here is that the authors, while offering a definition of risk that is more congruent with the mathematical definition of risk, again conflates the constructs of disruption and supply chain risk. In this dissertation, I will contend with Zsidisin's notion that offering a very refined definition of supply chain risk would indeed be counter-productive. However, I will emphasize the distinction between risk and disruption. Instead, I will offer a broader definition of the construct: a *supply chain risk* is the interaction between the total loss of an event if it were to occur and the likelihood of the event occurring, which could lead to a supply chain disruption. Note that this definition does not necessarily mean "expected loss". It could be operationalized from a Value-at-Risk perspective, among other perspectives. This definition is refined enough for theoretical contributions but not so refined where it limits other authors perspectives on the subject.

In this dissertation, I will assume that the probability of a disruption is itself a measurement of the culmination of all of a firm's risk mitigation strategies (as well as other non-risk-centric strategies). That is, the probability of a disruption is itself dependent on the agent's decision making. It is well known that agents make decisions

in the face of uncertainty by leveraging proper stochastic model design and optimization of those models (Prelec & Loewenstein, 1991; Bell, 1982). As a simple demonstration as to how managerial decision making affects risk, just consider the problem of safety stock:

If the demand D for a product in a single period of time follows a normal distribution $D \sim N(\mu, \sigma)$, and the firm would only like a 5% probability of running out of stock, then they would order Q units, where $P(D < Q) = .95$. In this example, the disruption would be "out of stock" and the probability of this is 0.05. This, however, was a consequence of ordering and holding the Q units (with a lot of other assumptions for this simple example).

A similar mathematical demonstration can be had with the notion of a second or backup supplier, among other risk mitigation strategies. Hence, I submit here that the probabilities of events that cause a disruption, that is, the likelihood of the risks manifesting, are partly dependent on the risk mitigation strategies employed by actors within the network. This is an important observation, as it allows for the ease of modeling the structure of supply networks and observing how this structure affects risk. It should be noted, however, that the likelihood of risk is not necessarily wholly determined by risk mitigation strategies alone.

There still could be other characteristics and behaviors of the agent that determine the probability of disruption. Since we have adopted (Carter et al., 2015) definition of the supply chain, it is also logical to assume that the probabilities will also change over time. Since supply networks are Complex Adaptive Systems (Carter et al., 2015), agents change their characteristics and behavior over time. Since the likelihood of risks manifesting into disruptions are partly dependent on these characteristics and behaviors, it is logical to conclude that the probabilities themselves are not static. In this dissertation, however, I will assume that the probabilities do not change over time since I will also assume that the network does not change over time. The inquiry of evolution of the networks is out of the scope of this dissertation. I do not currently seek to understand how risk evolves over time. Rather, I seek to understand how risk is *related* to network structure in the *static environment*.

1.1.3 Structural Network Characteristics and Supply Chain Risk

Not only did the previous examples demonstrate that risk and disruption are key aspects of supply chain management, but so too is structure. We saw that GM, Honda and Nisson had responded differently to the earthquake than did Toyota. We saw that Tesla took a different course of action in regards to their pricing than did Ford and Lincoln in response to the tariffs. This is due not only to the risk mitigation strategies that these firms had in place, but also due to their inherent structures that were designed.

Just as with the respective definitions of the supply chain, supply chain risk, and supply chain disruption, supply chain structure is also defined from a variety of differing perspectives in the extant literature. Choi contends that the "structure of a supply network can be viewed as the pattern of relationships among firms engaged in creating a sellable product" (Choi & Hong, 2002, pg. 470). In addition, he characterizes supply network structure into three primary dimensions: formalization, centralization, and complexity. This contrasts with Kim, who uses a categorization of different types of networks such as small world and scale-free networks, among others(Y. Kim, Chen, & Linderman, 2015; Watts & Strogatz, 1998).

Last, some authors have taken a more social-network perspective in combination with more traditional graph-theoretic measures such as density and distance (Carnovale & Yeniyurt, 2014; Carnovale, Yeniyurt, & Rogers, 2017). Concepts borrowed from social-network theory such as ego-networks, as well as traditional graph-theoretic measures such as eigenvector centrality, closeness centrality and betweenness centrality, have been used to determine their effect on Joint-Venture formation (Carnovale & Yeniyurt, 2014; Carnovale et al., 2017). Such a characterization of a network is, in this author's opinion, the most appropriate manner to characterize networks with respect to a pure structural perspective.

First, while defining structure based on dimensions does offer clarity for strategic reasons, current construct analysis of structure is heavily reliant on context. The same holds true for the prior attempts to classify networks into only a handful of categorizations. While conceptualization may offer clarify, classification greatly increases the

within-group variance in any response variables that may be considered within their connection to canonical network structure. (Carnovale & Yeniyurt, 2014) approach is the most appropriate for this dissertation's purpose as it does not risk the presence of the aforementioned issues. While still partly context-dependent, his use of the graph theoretic measures were used to properly characterize a network from a context-free, structural perspective. Hence, I will extend (Carnovale & Yeniyurt, 2014) approach by using additional graph-theoretic measures, and attempting to synthesize them into a handful of reasonable context-free constructs for network structure.

While (Choi & Hong, 2002) extended definition of supply chain structure involves the description of three dimensions, his original definition is actually quite fitting to the purpose at hand. Hence, for the purposes of this dissertation, I define the "supply network structure" in same more general sense that (Choi & Hong, 2002) first defined it prior to his discussion on three dimensions. More specifically, the "structure of a supply network can be viewed as the pattern of relationships among firms engaged in creating a sellable product" (Choi & Hong, 2002, pg. 470). This definition is general enough to encompass a large number of differing operationalizations without voiding any prior work on supply network structure.

1.2 Research Objectives and Inquiries

Now that I have clearly defined a few of the terms that are necessary for the primary inquiry of this dissertation, I shall offer two motivations for the primary topic at hand. First, I will give a formal justification for the motivation of the primary inquiry using the extant literature grounded within organization theory. Second, I will give a mathematical exposition for the primary problem at hand by demonstrating a small example. The mathematics used in this derivation will be explained in greater detail in the next chapter of this dissertation.

1.2.1 Conceptual Motivation

I will ground the inquiry for this dissertation in Resource Dependency Theory (RDT). Resource Dependency Theory states that the "survival of the organization is partially explained by the ability to cope with environmental contingencies" (Pfeffer & Salancik, 2003, pg. 258). Much of RDT is focused around the constructs of power and accessibility to resources. It states that firms require resources, and at times, access to those resources is controlled by other actors within an external environment. The level of control that these outside actors have on said resources establishes a power dynamic between two or more actors(Pfeffer & Salancik, 2003; Emerson, 1962; Casciaro & Piskorski, 2005). Indeed, Emerson notes that "it would appear that the power to control or influence the other resides in control over the things he values, which may range"(Emerson, 1962, pg. p32).

Throughout the process of navigating in this external environment of agents that have varying degrees of power, firms seek to establish relationships with others of lower levels of power and minimize or remain independent of relationships with high power firms. Indeed, "it is possible that when the buyer-supplier relationship dynamic is characterized by coercion, ... [a] supplier may simply wish to avoid attention, or at least has no inclination to help his customer beyond meeting basic performance expectatations" (Terpend & Ashenbaum, 2012, pg. 56).

Therefore, RDT considers connections between organizations as "instruments for reducing power imbalances and for managing mutual dependencies"(Drees & Heugens, 2013, pg. 4) which is further emphasized by (Casciaro & Piskorski, 2005). As a firm seeks to address these power imbalances, it determines which actors to connect with and which ones to avoid or minimize interaction. In doing so, they become connected or disconnected to a larger collection of firms, themselves having unique power dynamics. Indeed, "while a buyer-supplier relationship is a dyad, it is also part of networks that have come to bear on individual nodes to the relationship through each others extended business relationships" (Choi & Kim, 2008, pg. 6). Hence, it can be argued that the network structure of a supply network is the result of individual firms struggling to

address a variety of power imbalances.

While power is one aspect of RDT, dependence on resources is another. Recall that power is only one portion of Pfeffer's explanation of links between firms. He states that "an organization's vulnerability to extraorganizational influence is partly determined by the extent to which the organization has come to depend on certain types of exchanges for its operation"(Pfeffer & Salancik, 2003, pg. 46). He continues to describe two dimensions of the importance of the dependence: the magnitude and the criticality of the resource exchange. As such, he argues that any problems that arise out of the organization due to the level of importance that an organization places on a resource is due to the external environment. Namely, the ability to depend on, and access, the resource and power dynamics that are encountered through acquiring those resources.

Therefore, RDT explains that a supply network structure is formulated as a manifestation of power and the need to acquire resources via resource-exchanges (i.e. establishing links between firms). In addition, RDT can explain why firms within the supply network experience risk of disruption and the need to manage such uncertainties. These risks are inherent in the environment in which the firm operates. As Pffeffer states, "interdependence can create problems of uncertainty or unpredictability for the organization. This uncertainty ... derives from the lack of coordination of activities among social units"(Pfeffer & Salancik, 2003, pg. 42).

Part of the lack of coordination would be the lack of knowledge of restructuring of the network itself. Again, as Pffeffer indicates, "organizations facing uncertainty attempt to cope with it on occasion by restructuring their exchange relationships. The solution to one organization's uncertainties ... can create new uncertainties for other organizations"(Pfeffer & Salancik, 2003, pg. 42). This notion of agents within a network restructuring themselves to manage both uncertainty and power is consistent with (Carter et al., 2015) conceptualization of the supply network being a complex adaptive system (Carter et al., 2015).

As such, Resource Dependency Theory explains the following observations in relation to disruption risk and structural characteristics about a supply network:

1. Supply Network Structure is consistently in flux due to power imbalances, the need for survival via resource acquisition, and the need to reduce uncertainties.

2. As the Supply Network changes (due to power, resource, uncertainty), so too do the uncertainties of all members (and hence the chance of a disruption).

3. Structure and Risk of Disruption are endogenous. The structure, as suggested by RDT, itself creates uncertainties for actors within the network. As such, actors respond to the uncertainties by rearranging their structure. By the previous sentence, since the structure has changed, so too have the uncertainties and hence the risk for disruption. Again, actors respond by rearranging their structure, and so forth, ad infinitum. Therefore, the risk of disruption and supply network structure are intertwined and consistently in flux.

This dissertation seeks not to address all the observations made in the previous argument. Instead, it seeks to take the first step towards establishing the nature of the relationship between structure and risk. While the two are intertwined, they only evolve over the dimension of time. A good first step to understanding this evolution of the interaction between structure and risk is to first understand *how* structure and risk are related in the static sense. Given the aforementioned definition of structure, one can speculate that it has many possible dimensions that could be used to characterize, both conceptually and mathematically, the canonical network structure. Laying a foundation for which particular types of structure affect risk will lay a foundation for future work as to how these two constructs evolve with each other over time.

1.2.2 Mathematical Motivation for the Dissertation

While I have motivated the primary inquiry of this dissertation from a conceptual point of view, I would also like to do so from a mathematical point of view. Mathematically, the structure of a supply network is modeled as a directed graph (Nair & Vidal, 2011; Garvey et al., 2015; Carnovale & Yeniyurt, 2014). The nodes and arcs can represent different levels of analysis. If at the firm level, then nodes commonly refer to the firms within the network (relative to a fixed focal firm) and the arcs represent physical

material flow, information flow or a financial flow between firms within the network (Carter et al., 2015). The *graph* $G = (E, V)$ is then a combination of two mathematical sets: the edge set E and the node set V (Diestel, 2017).

As a result, many properties of this mathematical structure exist that help to explain the graph's "structure". Measures such as node count, edge count, degree centrality, eigenvector centrality, betweenness centrality, girth, connectedness, distance, and many others all allow for the scholar of graphs to understand the various structural properties, and even classifications (Diestel, 2017). Such properties are directly related to our primary inquiry at hand, and as such, I will leverage some of them in this motivation.

To motivate the topic of discussion mathematically, I will leverage two basic properties of graphs: distance and centrality. "Risk" is defined above as the interaction between the consequence of an event and the likelihood of the event. Here, and heretofore, I will measure the "consequence" of an event as the number of nodes within a network that are affected due to a disruption. The likelihoods of disruption are modeling using the concept of a *Bayesian Network*. I will forgo a formal discussion of this in the current motivation, and direct the reader to the next chapter for further clarification. For now, I will assume the reader is familiar with the fundamentals of conditional probability. The model of risk measurement of a supply network that I will employ is (Garvey et al., 2015) framework.

For the motivation, I will demonstrate the initial observations of relationships between the structural characteristics and the risk. Four networks of the same number of nodes, all of which are connected (i.e. any two nodes have a path between them), but with different levels of centrality and distance will be studied. Each of these structures are shown in Figure 1.1. From each network, I generate 99 *Bayesian Networks*, where all the conditional probabilities for each node in a single network are the same. That is, given a probability $0 < p < 1$, $P(X = 1|Parents(X)) = p$ and $P(X = 0|Parents(X)) = 1 - p$, where X is a node in the specific network generated.

I iterate the value of p starting at 0.01 and ending at 0.99, by every 0.01. Hence, for each of the four networks, I generate 99 Bayesian Networks (a unique Bayesian Network comprises of the structure and the collection of conditional probabilities). Hence, I will

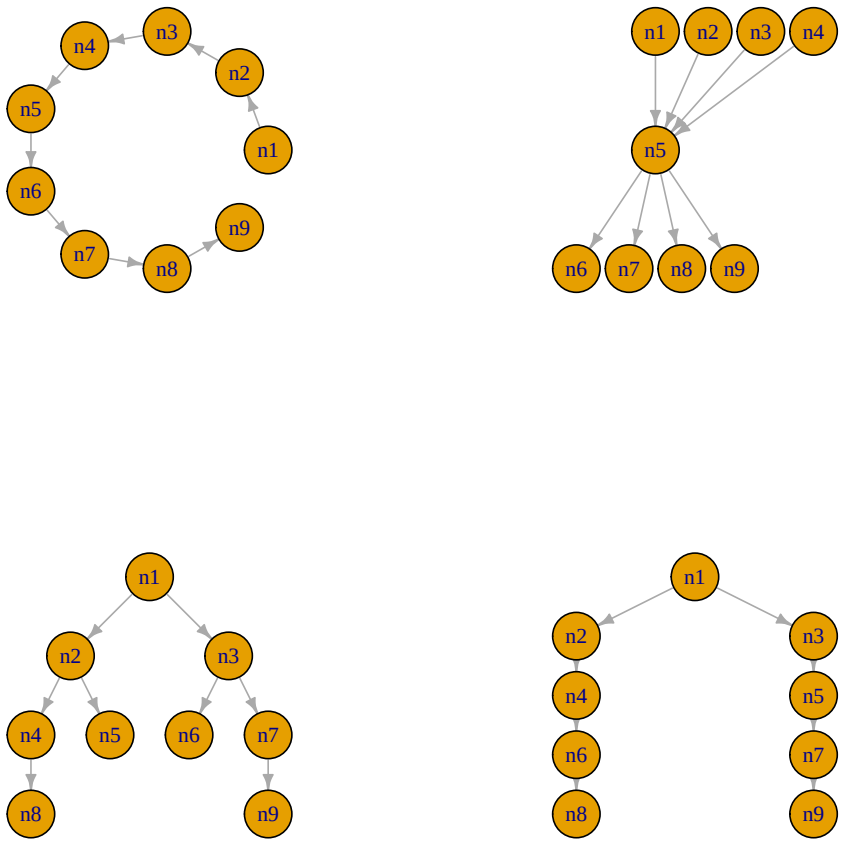


Figure 1.1: Four different directed networks, each one of which has 9 nodes.

study a total of 396 different Bayesian Networks. Varying the probabilities allows for enough heterogeneity in the probability data to be able to isolate the effects of structure on risk.

The code to generate these networks is in the appendix and all of the generation and analysis is conducted in R. I leverage the *igraph* and *bnlearn* packages in order to generate and analyze the Bayesian Networks. For this small demonstration, I analyze the generated data at two different levels of analysis: the network level and the node level. Arc analysis is ignored and reserved for the primary inquiry. Leveraging (Garvey et al., 2015) Risk Model, each node in the network represents a single risk (a generic risk of disruption) which holds a value of either 0 or 1 (the event occurred or the event did not occur). Risk is measured using the model's Expected Location Risk Contribution Factor (ELRCF) (Garvey et al., 2015). This is a node/arc-level risk metric that is calculated, for a node n, as:

$$ELRCF(n) = (1 + ERCF(n|n = 1))P(n = 1) + ERCF(n|n = 0)P(n = 0)$$

where ERCF is the Expected Risk Contribution Factor, which is defined for a node n assigned to a value $a \in \{0, 1\}$

$$ERCF(n = a) = E\left[\sum_{i \in D(n)} U(n_i)|n = a\right]$$

where $D(n)$ is the set of descendants of the node n and $U(n_i)$ is the cost if the event n_i were to occur (that is, it is the cost when $n_i = 1$). Here, as mentioned previously, I assume the consequence of a disruption is equal to the number of nodes where the disruption occurred. Hence, in this case, $U(n_i) = n_i \in \{0, 1\}$. The ELRCF for a node n measures the expected number of descendants of n that will experience a disruption due to a disruption (or no disruption) at the node n.

For the network-level analysis, the structural characteristics that I study in this demonstration are the radius, diameter, average betweenness centrality, average closeness centrality, average degree centrality and average eigenvector centrality. Overall "Network Risk" is calculated by finding the average ELRCF of all nodes, the standard deviation and the coefficient of variation $\left(\frac{\sigma_{ELRCF}}{\mu_{ELRCF}}\right)$. For the node-level analysis, the

structural characteristics that I study in this demonstration are betweenness centrality, closeness centrality, eigenvector centrality, and degree centrality. Risk is calculated at the node level by finding the ELRCF for the node.

In both levels of analysis, I control for the probability used to generate the data so that structure can be isolated. The simulation was run in R version 3.3 using the aforementioned packages and custom code on a Macbook Pro, OS X 10.9.5 with a 2.2 GHz Intel Core i7 and 8GB 1333 MHz DDR3 Memory. The *igraph* package was used to calculate the network characteristics and the *bnlearn* package was leveraged to represent the Bayesian Networks and to calculate the risk measures. Since the definition of the risk measures require a large number of numerical calculations, I leveraged the *bnlearn* package to conduct a Monte Carlo Simulation using the Liklihood Weighting Sampling algorithm, of which this package provides. I shall now illustrate some of the results from the initial demonstration.

First, I will illustrate the network-level analysis. While this analysis is not a focus of this dissertation, the same measures, calculations and analysis will be used to calculate k-ego network structural characteristics, and hence the demonstration illustrates the potential connections between these types of structures and risk. For the Network-Level analysis, we see from Figure 1.2 that both the distance and centrality measures have very little impact on the average risk in the network, but it could be argued to be slightly curvi-linear. However, this may be due to the homogeneity in the overall network structure itself (after all, I am only looking at 4 different network structures all of the same node and arc size). The same observations are made from the standard deviation of the risks and hence are not shown. However, upon further inspection of the data, plotting the coefficient of variation (CV) against the different structural variables seems to indicate an inverse quadratic relationship. Figure 1.3 indicates that the CV of risk starts small for smaller levels of centrality and distance, and increases to a precipitous. After this point, any increase in distance or centrality indicates a drop in the coefficient of variation for risk. What is interesting is that this observation holds for all centrality measures, despite eigenvector and betweenness centrality being inversely related to closeness centrality. If this association holds in my later analysis, then the

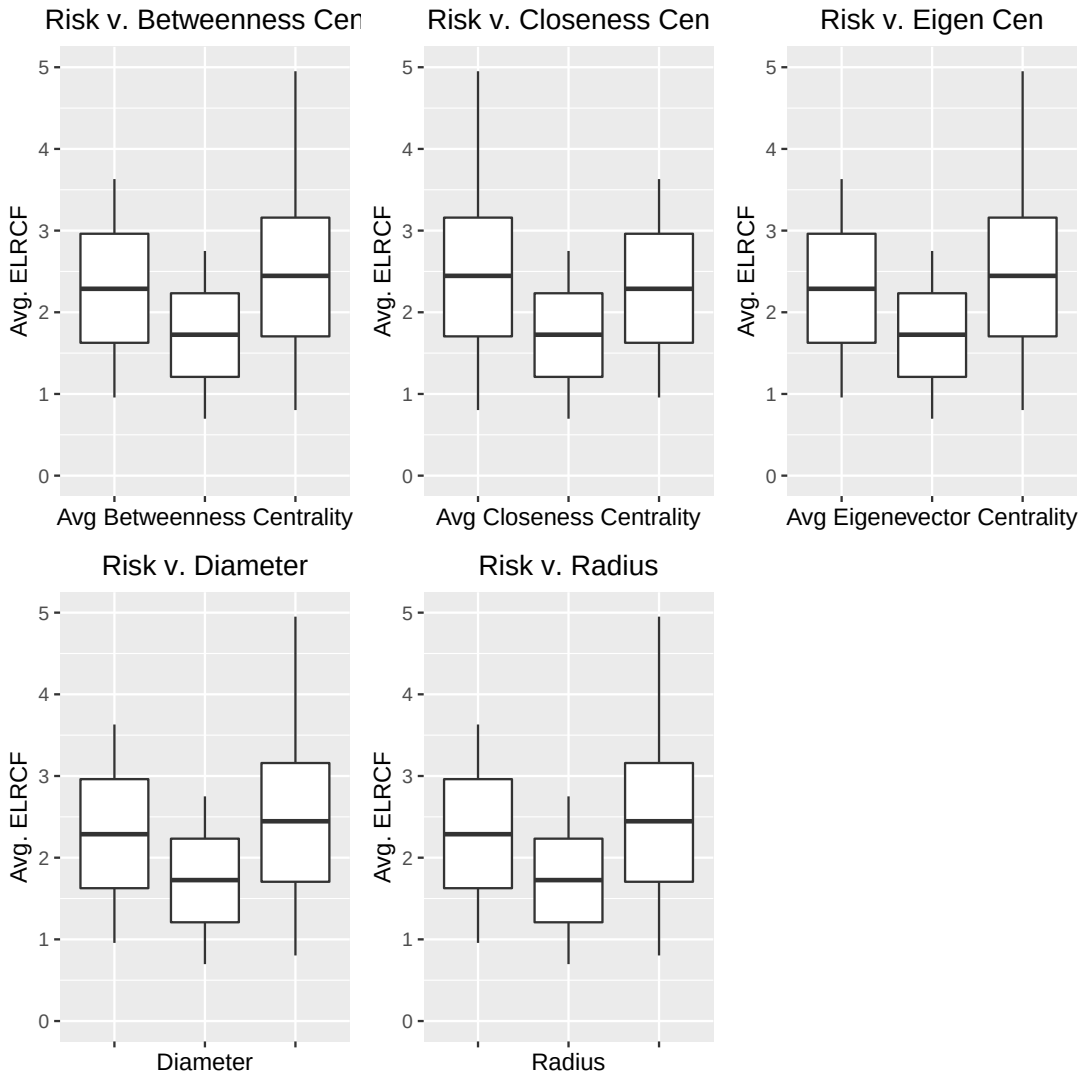


Figure 1.2: The Expected Location Risk Contribution Factor (ELRCF) plotted against the 4 different network's distance and centrality measures.

pattern that can be postulated is that regardless of the type of centrality, more central nodes are exposed to a maximum level of risk of disruption, which could be mitigated by creating additional connections beyond this level. The same would hold for network distance. These observations, however, may be due to the lack of sufficient structural data. However, despite the incredibly small sample size of varying structure, we do see that the CV of risk seems to be associated with the centrality and distance measures of network structure. We will see later if this holds true in the primary inquiry of this dissertation. I will now analyze the data from a node-level perspective. Figure 1.4 plots the risk of a node against it's different centrality measures. We can see from the plot that unlike at the network level, the node level seems to indicate a stronger association between the risk and centrality measures. Risk seems to be positively curvilinear with degree centrality. Generally, the risk seems to increase exponentially for the Eigenvector centrality. For Betweenness and Closeness Centrality, again, we notice that despite these being negatively associated with each other, each of them seems to exhibit a negative curvilinear association.

The visual analysis of the data simulated seems to indicate that risk, as measured by the expected number of nodes affected by a disruption, is associated with the varying dimensions of structural characteristics. The associations, at both the network level and node level, seem to indicate positive linear associations and possibly exponential and negative curvilinear. While this little demonstration did not have much heterogeneity within the structural data itself, it still serves as a motivation for the primary inquiry at hand.

Thus, from a theoretical perspective and empirical perspective, we can see that structure is possibly related with the risk within a supply network, *ceteris paribus*. This section has served as the foundation for the motivation of the current dissertation at hand. Now that the primary inquiry is supported by the prior motivation, I will summarize the primary research objectives of this dissertation.

As the motivation has indicated, the need to characterize the relationships between the static structure of a network and risk is important to understand it's evolution over time. The motivation also argues that understanding this evolution is important for

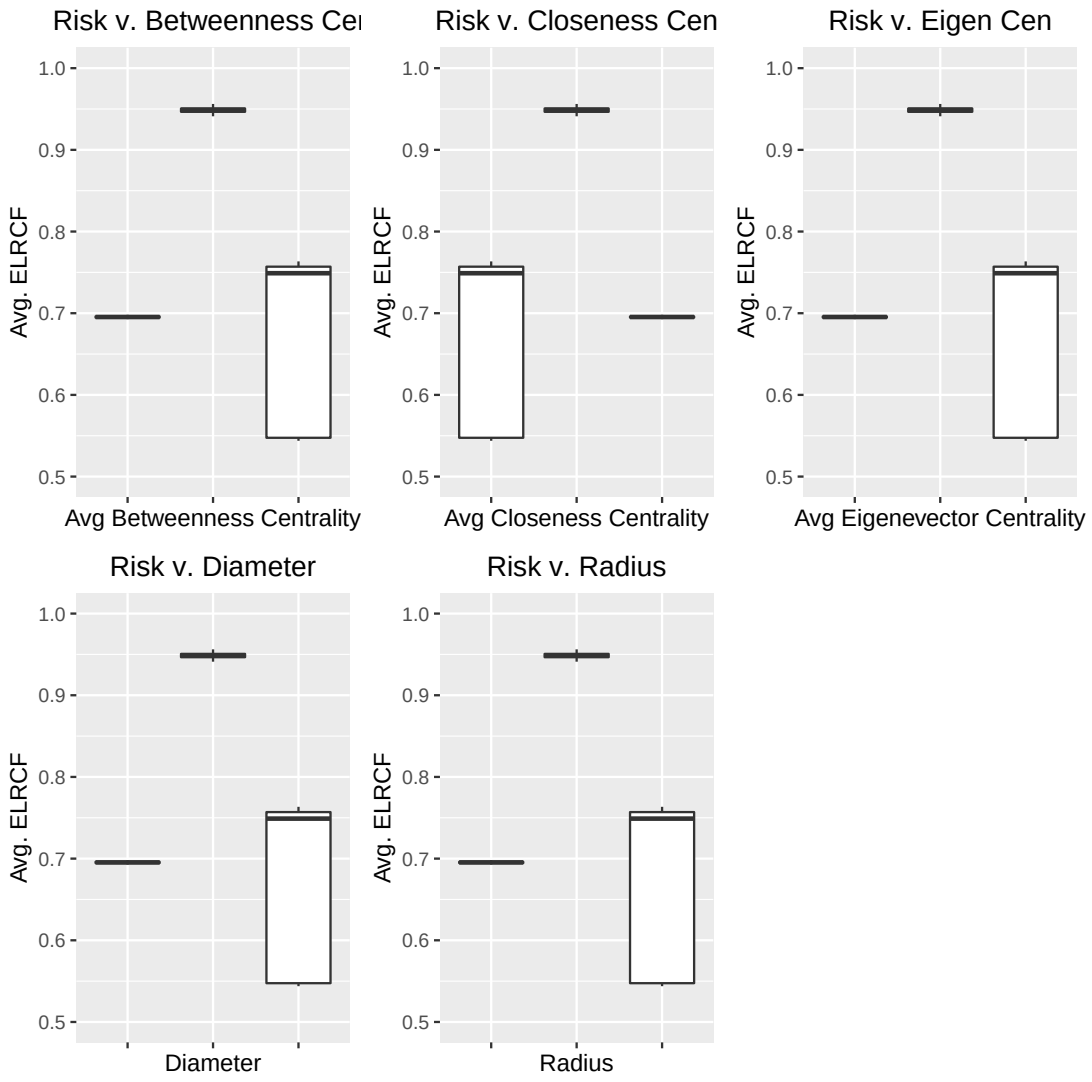


Figure 1.3: The Coefficient of Variation of the Expected Location Risk Contribution Factor (ELRCF) plotted against the 4 different network's distance and centrality measures.

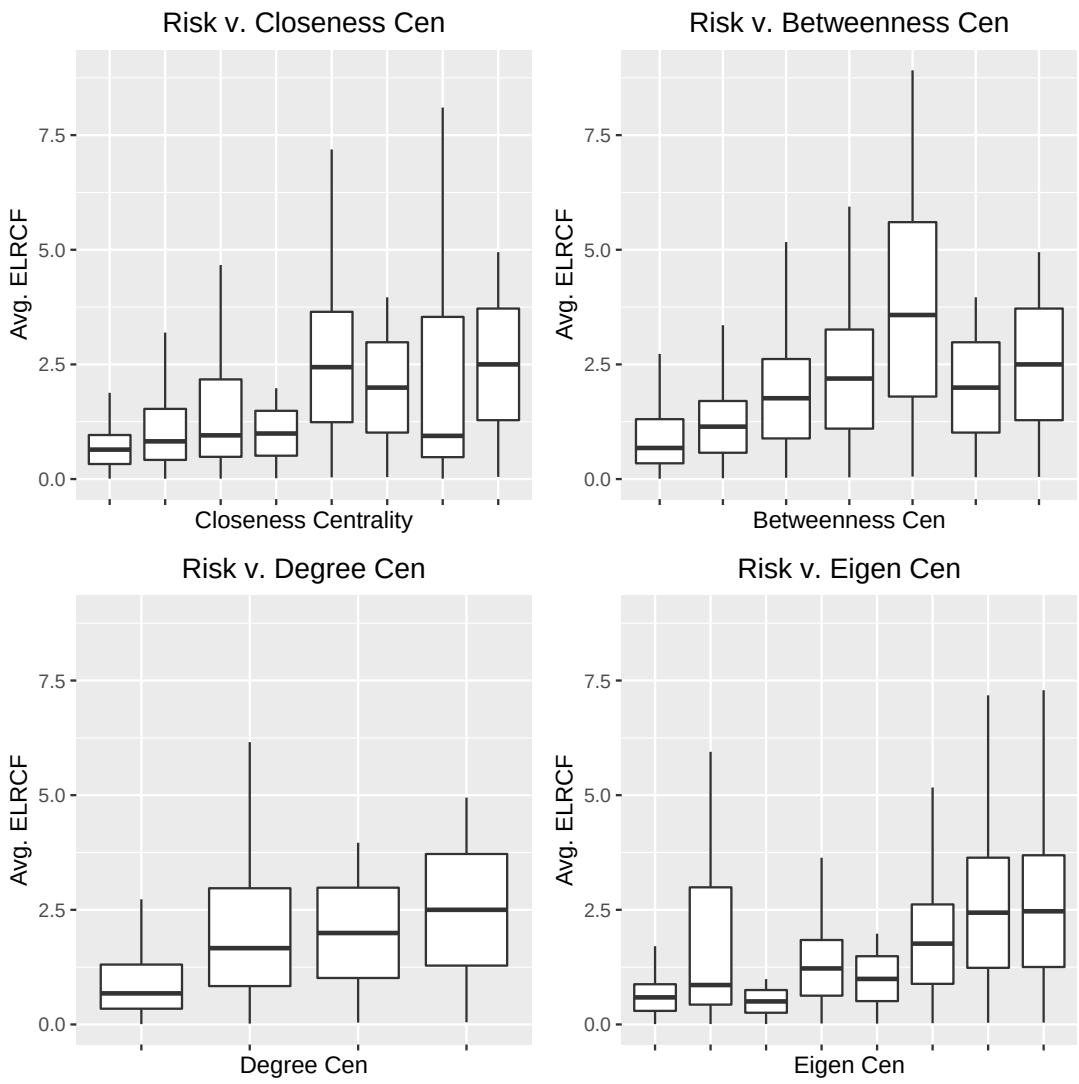


Figure 1.4: Node-Level Risk plotted against the node-level centrality measures.

firms that seek to survive. Again, according to RDT, firms survive by being able to appropriately respond to environmental changes. Since the structure of the network is inherently canonical of the external environment, the firm therefore must understand the association to design effective risk mitigation strategies.

In the mathematical motivation, I demonstrated that there seems to be some evidence that warrants further investigation into this association. We had seen that at the network level, structure impacts the coefficient of variation of risk, while at the node level, it impacts the risk itself. Since firms are embedded in larger structures (Nair, Blome, Choi, & Lee, 2018; Choi & Kim, 2008), it would be worth while to determine if structures at a higher level of analysis affect risk at these levels of analysis (for example, does ego-network structure affect ego-network level risk?) or, to determine if the risk at a higher level of analysis affects risk at lower levels (for example, does ego-network risk affect node/arc level risk?). As I will demonstrate in a later chapter, networks have different levels of analysis of which can be effectively described with the notion of a k-ego network. In this dissertation, I seek to understand how the structure of the ego network of the ego network of a firm affects not only the lower ego network risk of a firm but also the firm risk itself. In addition, understanding arc-level risk would serve useful if it is found that risks are affected differently for arcs than for nodes. Therefore, I submit the following objectives that this dissertation seeks to complete:

1. Demonstrate the definition of different levels of embeddedness and visibility through the construct of a k-ego network.

2. Demonstrate how structure differs across the k-ego networks.

3. Determine the constructs of k-ego network structure that are directly predictive of k-ego network risk.

4. Determine the nature of the relationship between k-ego networks structure and k-ego network's risk for $k = 0, 1, 2$.

5. Determine the nature of the relationship between structure and arc/node risk.

6. Determine the relation between higher k-ego-network levels of risk and lower k-ego-network levels of risk.

Chapter 2

Literature Review

2.1 Supply Chain Risk and Disruption

2.1.1 Supply Chain Disruptions

A *supply chain disruption* is an "unplanned and unanticipated event that disrupts the normal flow of goods and materials within a supply chain" (Craighead et al., 2007;P. R. Kleindorfer & Saad, 2005; Hendricks & Singhal, 2005b). Special attention has been given to supply chain disruptions in the extant literature due to the rise of global sourcing, which inherently increases the potential for disruptions to occur, and due to the movement towards optimizing the characteristics of the supply chain such as responsiveness, agility and inventory (Blackhurst*, Craighead, Elkins, & Handfield, 2005). In addition, longer paths and shorter clock speeds lead to increased opportunities for disruption (P. R. Kleindorfer & Saad, 2005). Not only has attention to disruptions been given by academics, but also by practitioners, of whom "consider supply chain disruptions and their associated operational and financial risks to be their single most pressing concern"(Craighead et al., 2007, pg. 149;Green, 2004).

Supply Chain Disruptions can have significant impact on a firm's performance. Indeed, "a disruption affecting an entity anywhere in the supply chain can have a direct effect on a corporations ability to continue operations, get finished goods to market or provide critical services to customers" (Jüttner, Peck, & Christopher, 2003). Much of the extant literature has studied a variety of detrimental impacts that a disruption within the network has on a firm. For example, Duhhadway had suggested that the impact on performance is dependent on the type of disruption (DuHadway et al., 2017) and Rosales indicated that supply chain disruptions such as out of stock (OOS) can

contribute a negative role in retailer and manufacturing performance (Rosales, Whipple, & Blackhurst, 2018). (Hendricks & Singhal, 2005a) found that supply chain disruptions can lead to a decrease in profitability and net sales, as well as an increase in costs and negative inventory performance (Hendricks & Singhal, 2005a).

In addition, supply chain disruptions can be problematic for production scheduling (Bozarth, Warsing, Flynn, & Flynn, 2009) as well as other activities of the firm. It has also been shown that disruptions originating from transportation activities can negatively impact inventory levels, goods in transit as well as the number of unfilled customer orders (Wilson, 2007). Others have found that speed to recovery lies within the nature of the disruption itself, as well as having an impact on financial and service performance (Macdonald & Corsi, 2013). Other firm characteristics, such as robustness and agility, have been demonstrated to amplify/depress the result of disruptions (Wieland & Marcus Wallenburg, 2012). In addition, supply chain disruptions can have an impact on a firm's competitive performance, schedule attainment and customer satisfaction (L. Zhao, Huo, Sun, & Zhao, 2013).

As the aforementioned anecdotes illustrate, when a disruption occurs within a supply network, firms often have the need to respond with action. More generally, it has been suggested that firms react to disruptions in one of three ways: parametrical, process or structural (Ivanov, Dolgui, Sokolov, & Ivanova, 2017). In the case of parametrical, firms choose to adjust important parameters such as inventory, lead time or prices (Ivanov, Dolgui, Sokolov, & Ivanova, 2017; Bugerta & Lascha, 2018). On the other hand, firms may choose to respond by adjusting their processes or their structural characteristics within the network. Prior research has suggested that firms often form expectations of subsequent changes to their external environment by other firms in the network. Indeed, when a disruption occurs, firms speculate on the continuity of a dyadic relationship. Such speculations involve firm perceptions of opportunism, dependence on opportunism, and relationship continuance (Ro, Su, & Chen, 2016). Such perceptions are inherently emotional, behavioral, and subjective, and hence lead firms to decide if they should retain their supplier relationships (Polyviou, Rungtusanatham, Reczek, & Knemeyer, 2018). Others in the extant literature have suggested that firms could

respond through adjustments to price, where is has been shown that "significant cost savings for the supply chain" can be achieved (Bugerta & Lascha, 2018, pg. 33).

Yet another form of response is to implement a disruption and risk management system within the organization(Craighead et al., 2007; DuHadway et al., 2017). An important requirement in order to design a proper risk and disruption mitigation and response system is to first understand the different types of disruptions as well as the common drivers of them. As Duhadway noted, "in the supply chain literature, disruptions are often treated as homogenous regardless of the nature of the disruption"(DuHadway et al., 2017, pg. 4). It should be emphasized here, however, that "supply chain risk", which is considered to be an event that "causes" a disruption, and "supply chain disruption" are different in regards to their nature. Despite this, it is still important to have a classification of types of disruptions as well as an understanding of how they manifest. Duhadway has suggested that disruptions can be classified along two primary dimensions: Behavioral Intent (Intentional/Unintentional) and Disruption Location (Endogenous/Exogenous) (DuHadway et al., 2017). Such a classification allows the supply chain manager to better understand the potential impact on firm performance due to the different types of disruptions.

Prior literature has made a few findings as to what would drive a disruption. Firm behaviors such as risk taking propensity, security compliance measure implementation and safety stock have been found to affect the frequency of supply chain disruptions (Park, Min, & Min, 2016). Other actors within the environment in which the firm operates can also affect the occurrence of disruptions, such as violent cargo theft (Ekwall & Lantz, 2018; DuHadway et al., 2017). In addition, a firm's willingness to embrace Supply Chain Disruption Management and Supply Chain Risk Management affects the frequency in which disruptions manifest in the supply network (Revilla & Saenz, 2017).

Part of proper disruption management involves the ability to detect a disruption. Many methodologies have been suggested for this purpose. Craighead defined a *severe* node as one where the disruption at that node affects a large number of other nodes. Hence, identifying severe nodes ahead of time would help in the detection of disruptions (Craighead et al., 2007). Decision support systems that involve the gathering of data

and scenario analysis is another solution to the problem of detection(K. Zhao, Scheibe, Blackhurst, & Kumar, 2018). Other methods include Petri Net models (Blackhurst, Rungtusanatham, Scheibe, & Ambulkar, 2018) and the use of other forms of decision support systems (T. Wu, Blackhurst, & Ogrady, 2007). A more recent trend has been to consider the aspects of uncertainty (i.e. supply chain risk) into the detection of disruptions leveraging Bayesian Networks(Garvey et al., 2015).

Not only has detection been a concern for academics and practitioners, but so too has been the design and implementation of strategies in order to mitigate against said disruptions. This stream of literature is vast. Some of the recent work conducted on the matter has involved the analysis of strategies that involve the use of primary suppliers, pre-positioned warehouses, backup suppliers and protected suppliers (Kamalahmadi & Parast, 2017). Others involve the creation of an inference system to predict a disruption in regards to demand and inventory as well as a reactive strategy by adjusting production levels in response (Paul, Sarker, & Essam, 2017).

Another mechanism to mitigate disruptions is by the creation of a supplier portfolio, where different suppliers are used for different purposes and frequencies in the wake of a disruption (Sawik, 2017; Namdar, Li, Sawhney, & Pradhan, 2018; Sawik, 2018a, 2018b). The use of order-up-to policies in inventory systems have also been shown to help mitigate disruptions as well (Schmitt, Kumar, Stecke, Glover, & Ehlen, 2017). Even more mechanical-based decision making tools, based on Agent-Based models, have been suggested as an effective way to mitigate(Blos, da Silva, & Wee, 2018), where human decision making is delegated to an automated decision support system.

Another aspect of disruption mitigation is within the design of the network structure itself. Indeed, many in the extant literature have studied how best to design a network when disruption is a possibility. Many of the designs are based on optimizing a variety of different costs pertaining to inventory, routing, capacity, operating, ordering, as well as more supply chain centric parameters such as income, responsiveness, flexibility, environmental impacts, social responsibility, and risk/robustness metrics (Govindan, Fattahi, & Keyvanshokooh, 2017). Other network designs are based on incorporating multiple sourcing, extra capacity and emergency stock strategies while taking into

account resiliency and competitiveness (Rezapour, Farahani, & Pourakbar, 2017; Margolis, Sullivan, Mason, & Magagnotti, 2018; Sáenz, Revilla, & Acero, 2018). Some network designs are not designs at all, but rather they are the manifestation of decision making, such as facility location, due to disruptions (Jabbarzadeh, Fahimnia, & Sabouhi, 2018; Jabbarzadeh, Haughton, & Khosrojerdi, 2018).

In the literature, there have been multiple attempts to determine optimal decision making by employing operations research models. Such attempts include optimization of service level (Sawik, 2016), return flows (Ivanov, Pavlov, Pavlov, & Sokolov, 2017), pricing and inventory decisions (Huang, He, & Li, 2018), budget allocation (Zhang, Zhao, & Pang, 2018), and supply chain performance as measured by supply chain system profit (Giri & Sarker, 2017). The literature in this area is quite large, and so I direct the reader to a recent review in this body of work (Snyder et al., 2016).

2.1.2 Supply Chain Risk and Risk Management

The Need for Supply Chain Risk Management (SCRM)

As previously mentioned, a supply chain risk is distinct from a supply chain disruption. Indeed, "it is worth noting the distinction between disruption and risk. Whereas disruptions are typically manifestations of supply chain risks; risk can be present without a disruption occurring" (DuHadway et al., 2017, pg. 2). Therefore, we can conceptualize a supply chain risk as an event, which holds a certain level of consequence and a likelihood of occurrence, that *could* lead to a supply chain disruption (G. A. Zsidisin, 2003; Pettit, Fiksel, & Croxton, 2010; Craighead et al., 2007). As I noted earlier, we can define supply chain risk as an interaction between an adverse event's likelihood and consequence that can manifest into a supply chain disruption. Such a definition allows for a wide array of operationalizations and different perspectives on "risk".

Supply chain risks are as important to understand and mitigate as supply chain disruptions. Previous literature has suggested that supply and demand risk both affect supply chain performance (Wagner & Bode, 2008). In addition, supply chain risks have led managers to feel the need, driven by internal and external factors, to change

the structure and relationships of these networks (Ritchie & Brindley, 2007). These changes, along with supply chain risk itself, could indeed lead to a reduction in financial performance depending on the resulting structure (Carnovale & Yeniyurt, 2015; Manuj, Esper, & Stank, 2014; Kilubi & Haasis, 2015; Kordlouie, Sadeghi, & Sadeghi, 2018). It also has been demonstrated that a variety of supply chain risks affect SCOR-Based Performance Metrics (Abolghasemi, Khodakarami, & Tehranifard, 2015; Truong Quang & Hara, 2018), innovation performance (Wiengarten, Humphreys, Gimenez, & McIvor, 2016; Kwak, Seo, & Mason, 2018), strategic performance (Revilla & Saenz, 2017), and operational performance (Salam, Ali, & Kan, 2017).

While supply chain risk is inherent within supply networks and can lead to detrimental effects on performance, researchers have found that the dynamics of risk perception by practitioners is an important consideration that must be taken into account. In a survey that Zsidisin conducted on managers within manufacturing firms, it was found that even "though the chance exists of an undesirable event occurring that would halt operations, the time and resources necessary for purchasing to plan ahead was not perceived as being important for several of the firms" (G. A. Zsidisin, Panelli, & Upton, 2000, pg. 8). However, when firms do perceive high levels of supply risk within their supply base, the frequency of disruptions tend to be higher. If firms recognize these risks and take the time to attempt to mitigate them, however, then they can achieve a reduction in disruption frequency and from decreased performance (G. A. Zsidisin & Wagner, 2010; Brusset & Teller, 2017; Kurniawan, Zailani, Iranmanesh, & Rajagopal, 2017; Adebola Adeseun, Anosike, Garza-Reyes, & Al-Talib, 2018).

Therefore, risk mitigation, and hence disruption frequency and performance, is inherently dependent not only on higher levels of analysis of the organization, but also to lower levels. That is, the state of these constructs are inherently dependent on individual emotion of actors within the organization. For example, researchers have found that risk perceptions vary across the different levels of the organization, despite managerial experience having no impact on such perceptions (Adebola Adeseun et al., 2018).

At the supply chain level, it has also been found that the vulnerability of a supply chain is partly dependent on a firm's perception of time and dependence of a relationship

between two firms (Svensson, 2002). Recently it was found that despite having success in risk mitigation strategies due to the perception of risk, the communication of such success actually had an inverse effect on the continence of those strategies (DuHadway, Carnovale, & Kannan, 2018). Therefore, the dynamics of risk management and perception are much more complicated than tackling risks from a pure strategic point of view. More specifically, not only are risks of a concern for firms within a supply network, but so too are the various behavioral dynamics of actors across multiple levels of the supply network and the organization.

The effects that risk can have on a firm within a supply network, as well as the behavioral dynamics of actors within the network and the organization, are important for practitioners to consider while crafting their strategies to meet their objectives. Simply stating that managers should "account for risk", however, is vague and ambiguous. Early researchers in the field had identified that supply chain risk is multifaceted and requires attention to detail. For example, Ellram had identified that when firms engage in relationships, there exist risks for both parties along economic, managerial and strategic dimensions (Ellram & Cooper, 1990).

In addition, if firms fail to consider the flexibility of those relationships, then any risks to and from the supply chain that are inherent within them will be difficult to change at a later point in time, and hence, a detailed risk analysis is necessary before any establishment or other changes to a relationship are made (Porter, 1991; Swaminathan, Smith, & Sadeh, 1998). Furthermore, decisions that firms make within a supply network are also embedded in a more general environment of "the world", or as Pfeffer defines, the "external environment" (Pfeffer & Salancik, 2003).

Risks are not only inherent within the decisions that a firm implements, but also within this external environment (Barry, 2004; Giunipero & Aly Eltantawy, 2004). The risks that are present, mutually exclusive to decisions within supply network, but rather within the external environment in which the supply network operates, indeed is an important consideration for firms (Paulraj & Chen, 2007; Pfeffer & Salancik, 2003). These risks are complex in nature, and can lead to a chain of events (i.e. a chain of other risks), that eventually "trickle down" to risks within the supply chain

environment (Garvey et al., 2015; Qazi, Quigley, & Dickson, 2015).

Firms could address the risks and the manifestation of these risks (i.e. disruptions) through either proactive or reactive strategies (Elluru, Gupta, Kaur, & Singh, 2017). However, if firms seek to fulfill their performance objectives, or seek to fulfill a greater goal of successful integration, they must address the threat of supply chain risks within a more organized and structured approach (Carter & Rogers, 2008; Flynn, Koufteros, & Lu, 2016; Blome, Schoenherr, & Rexhausen, 2013). Indeed, as "uncertainty in the business environment continues to increase, organizations need to adopt a more systematic and structured approach to supply chain risk management" (Cristopher, 1992, pg. 8).

SCRM Frameworks: Evolution and Current State

The extant literature currently has differing views on the definition of *Supply Chain Risk Management (SCRM).* For example, "an inter-organisational collaborative endeavour utilising quantitative and qualitative risk management methodologies to identify, evaluate, mitigate and monitor unexpected macro and micro level events or conditions, which might adversely impact any part of a supply chain" (W. Ho, Zheng, Yildiz, & Talluri, 2015, pg. 5036), while Chopra had seen SCRM as more of a strategic activity of which entailed stress testing and subsequent tailoring of risk-mitigation strategy to an individual firm (Chopra & Sodhi, 2004). While much debate still ensures, what seems to be agreed upon by most academics is that it lies at the intersection of Supply Chain Management and Risk Management (Khojasteh-Ghamari & Irohara, 2018).

One of the earliest use of the term SCRM was by Christopher, where he argued that risks need to be managed in a structured and systematic way by having a team of individuals "audit risk across the supply chain and ... develop and implement strategies for the mitigation of identified risk" (Cristopher, 1992, pg. 9). In parallel with some in the literature calling for a more structured SCRM framework, risk management was also being viewed as a strategic management activity, where "risk" was being understood from the perspective of loss of competitive advantage (Lonsdale, 1999). Researchers soon began to realize the need for a well defined taxonomy of risks that are present

within the supply network and the need for a framework to manage these risks.

Harland was one of the first authors to materialize Christopher's notion of "audit risk", where he proposed a process of SCRM by (1) mapping the supply network, (2) identifying risks by their type and their potential loss, (3) assessing the risk by understanding it's likelihood, exposure, triggers and loss, (4) managing the risk by developing risk positions (i.e. reactive, defensive, proactive or analytical) and scenarios, (5) form a collaborative supply network risk strategy, and (6) implement the supply network risk strategy. He was among the first of many authors to suggest that SCRM is indeed a collaborative effort among many actors within the network, but had noted that "their own view of how to deal with risk in the network may not coincide with other actors and is likely to give rise to destabilising of the network at some point" (C. Harland, Brenchley, & Walker, 2003, pg. 56).

In addition to Harland's characterization of SCRM, Zsidisin had proposed a framework for handling supply risk (which, while a consideration for SCRM, is itself distinct from supply chain risk). His framework involved applying Agency Theory to construct a method for managing supply risk. He proposed that first, supply risk sources should be identified and analyzed (with both likelihood and potential loss as factors for consideration), and then a Behavior-Based or Buffer-Based form of management should be employed (G. Zsidisin & Ellram, 2003). He later refined his framework of supply risk by building off of Hardland's framework, and argued that his "view of supply risk becomes a cornerstone for understanding supply chain or supply network risk, which would involve defining and studying risk at numerous inter- and intra-organizational levels" (G. A. Zsidisin, 2003, pg. 21).

Christopher and Peck emphasized the importance of collaboration between members of a supply network as a key component of SCRM. They stressed that knowledge sharing is important towards the ultimate goal of risk reduction. Building on the assumption that "collaborative working in the supply chain is that the exchange of information can reduce uncertainty" (Christopher & Peck, 2004, pg. 17), they argue that knowledge sharing via collaborative planning is a cornerstone of SCRM, where "knowledge" can be broken into three primary levels: strategic, tactical and operational. In addition,

they place SCRM within a larger context of resiliency, and are among the first in the literature to introduce the constructs of supply chain vulnerability and robustness to supply chain risk.

In their landmark paper, Kleindorfer and Saad were among the first to distinguish between a supply chain disruption and a supply chain risk. They argued that risk emanates from two sources: (1) issues manifesting from coordination of supply and demand and (2) disruptions to normal activities. They argued that the latter source itself comprises of a large array of different sources of risk that could be either internal or external to the firm or supply chain and requires it's own process. Their proposed process for managing these disruptions include Specifying the Risk Sources and Vulnerabilities (S), Assessing the Risks (A), and Creating Risk Mitigation (M), of which is analogous to Harland's conceptualization of SCRM. In addition, they propose a set of ten principles for achieving these three objectives. (P. R. Kleindorfer & Saad, 2005). Their work has served as the template of a SCRM model that years of subsequent and diverse research streams have been rooted within (Cohen & Kunreuther, 2007; Braunscheidel & Suresh, 2018).

These principles were later, yet separately, echoed by Handfield and Blackhurst and summarized through their definition of *Supply Chain Risk Management* as being the "integration and management of organizations within a supply chain to minimize risk and reduce the likelihood of disruptions through cooperative organizational relationships, effective business processes, and high levels of information sharing" (Handfield, Blackhurst, Elkins, & Craighead, 2007, pg. 30). Many authors since then have contributed individual advancements along each of the original dimensions proposed in the frameworks proposed by Hardland, Kleindorfer, and Handfield, respectively (G. A. Zsidisin & Ritchie, 2009; Narasimhan & Talluri, 2009; W. Ho et al., 2015; Fahimnia, Tang, Davarzani, & Sarkis, 2015; Fan & Stevenson, 2018; Heckmann, Comes, & Nickel, 2015).

However, these original frameworks were far from fixed, and new SCRM frameworks were continually being proposed in the literature as new areas of Supply Chain Management were discovered and developed. Along with the expansion of new topics and constructs within SCM, SCRM frameworks required updating to include newer

Topic	Authors
Supply Chain Structure	(Ritchie & Brindley, 2007; Lavastre, Gunasekaran, & Spalanzani, 2014) (G. Li, Fan, Lee, & Cheng, 2015)
Performance Outcomes	(Ritchie & Brindley, 2007; Manuj & Mentzer, 2008) (DuHadway et al., 2017; G. Li et al., 2015)
Risk/Driver Classification	(Manuj & Mentzer, 2008; Pfohl, Köhler, & Thomas, 2010) (Lavastre et al., 2014)
Behavior and Perception	(Manuj & Mentzer, 2008; Lavastre et al., 2014)
Assessment and Measurement	(Ritchie & Brindley, 2007; Manuj & Mentzer, 2008) (Tummala & Schoenherr, 2011; Klibi & Martel, 2012) (Blos, Hoeflich, & Miyagi, 2015; Garvey et al., 2015; Pettit et al., 2010)
Decision Modeling	(Manuj & Mentzer, 2008; R. S. Gaonkar & Viswanadham, 2007) (Klibi & Martel, 2012)
Mitigation Strategy	(Manuj & Mentzer, 2008; Pettit et al., 2010) (Tummala & Schoenherr, 2011)
SCRM Performance	(Manuj & Mentzer, 2008; Agrogiannis & Kinias, 2018) (Khan, Babar, Tareen, Naeem, & Gailani, 2018)
Strategy Selection	(Manuj & Mentzer, 2008)
Coordination & Collaboration	(Arshinder, Kanda, & Deshmukh, 2011; G. Li et al., 2015)
Resilience	(Pettit et al., 2010)
SCRM Implementation	(Pfohl et al., 2010)
Information Sharing	(Pfohl et al., 2010; G. Li et al., 2015; DuHadway et al., 2017)
Firm Characteristics	(Lavastre et al., 2014; G. Li et al., 2015)
SC Integration	(DuHadway et al., 2017)
SC Visibility	(DuHadway et al., 2017)
SC Sustainability	(Agrogiannis & Kinias, 2018; Khan et al., 2018)

Table 2.1: SCRM Framework Dimensions

dimensions to account for the change in SCM research. Hence, subsequent frameworks were proposed, but kept the spirit of the aforementioned principles while incorporating new dimensions and elaborations on old ones. These both include supply chain structure (node vs chain level), performance outcomes, additional classification of risks and risk drivers, involvement of actor behavior and risk perception, risk measurement and assessment, decision modeling, mitigation strategy, risk management performance measurement, strategy selection, coordination/collaboration tools and strategies, resilience, implementation, information sharing, firm characteristics, supply chain integration, visibility, and sustainability. Table 2.1 summarizes a sample of the literature in each of these respective dimensions of new frameworks proposed within the past two decades.

Risk Taxonomies: A Lack of Consensus

Now that I have offered a summarization of Supply Chain Risk Management frameworks, I feel it necessary to elaborate on only a handful of the aforementioned dimensions. In this dissertation, I shall avoid elaboration on a discussion of all of these

dimensions, as not all are relevant, directly or indirectly, to the topic at hand. However, I will elaborate on risk categorization, identification, assessment, measurement, mitigation strategy, and network design, whilst addressing some common problems that SCRM seeks to solve, as I feel that these are relevant to the topic at hand and offer a more holistic context to the primary inquiry of the dissertation.

As mentioned, many frameworks have proposed a categorization of risk (Ellram & Cooper, 1990; Manuj & Mentzer, 2008; Pfohl et al., 2010; Lavastre et al., 2014). Such a categorization has been shown to aid in the examination of possible consequences of said risk (C. Harland et al., 2003). In addition, Kleindorfer and Saad has argued that traditional "risk management is ... undertaken for each key process to identify vulnerabilities, triggers for these vulnerabilities, likelihood of occurrence, and mitigation and risk transfer activities" that give rise to "a taxonomy of major disruption risk categories that confront key assets of the firm" (P. R. Kleindorfer & Saad, 2005, pg. 58). That is, they argue that classification is not necessarily a cause of risk management but rather a manifestation of such an activity.

Classification of supply chain risk in the extant literature is vast, spans across multiple and sometimes conflicting dimensions, yet serves as a useful tool within both the risk measurement and management processes. Measurement of risk is inherent in the mitigation strategies, external environment, supply network environment and firm characteristics. Therefore, classification of risk is a necessity for proper estimation of risk measurement (via loss and probability of disruption measurement).

Some of the earliest mentions of the need for firms to categorize supply chain risk dates back to Porter and Ellram, respectively. An observation Porter made regarding the literature at that time was that "taxonomies have begun to emerge which attempt to categorize the ways in which firms can respond to uncertainty" (Porter, 1991, pg. 107). Yet another early contributor to this stream of literature was Ellram, who offered one of the first supply chain risk classification systems, proposing that firms, when they engage in relationships, put themselves "at risk" along the categories of economic, strategic and managerial (Ellram & Cooper, 1990). Soon thereafter, other authors began to offer their own versions of a risk classification system.

Giunipero had suggested a breakdown of common "risk creators", which included material availability, long distances, insufficient capacity, demand fluctuations, technological changes, financial instability, labor instability, and management turnover (Giunipero & Aly Eltantawy, 2004). Harland had designed a more general classification of risk from the literature at that time, where he created a taxonomy that included strategic, operations, supply, customer, asset impairment, competitive, reputation, financial, fiscal, regulatory, and legal risk (C. Harland et al., 2003). Gaonkar opted instead to create a context-free classification, where risk was categorized in three different groups: deviation (a difference from an expected operating level), disruption (the SC network is altered), and disaster (when the external environment is altered)(R. Gaonkar & Viswanadham, 2004).

Cavinato took a more structure-based approach to risk classification by offering five levels of which risk can embed itself: physical, financial, informational, relational, and innovational (Cavinato, 2004). In their landmark paper, Chopra and Sodhi suggested that risk should be categorized as disruptions, delays, systems, forecast, intellectual property, procurement, receivables, inventory, and capacity (Chopra & Sodhi, 2004). Rao had synthesized the extant literature on risk typologies into five categories: environmental, industry, organization, problem-specific, and decision maker risk. Subsequently, each of these categories had subcategories (Rao & Goldsby, 2009). At the same time, Wagner and Bode suggested a different typology that was more analogous to that of Chopra and Sodhi. They argued that risk should be classified by demand side, supply side, regulatory, legal, bureaucratic, infrastructure, and catastrophic risk (Wagner & Bode, 2009).

Monroe's survey of the literature suggested that he found 39 sources of risk in the supply chain literature, suggesting that academics have yet to agree on a synthesized taxonomy of supply network risks (Monroe, Teets, & Martin, 2014). In an attempt to address this problem by synthesizing, forming a consensus, and standardizing a risk-classification system, Rangel proposed, via a content-analysis on the literature, a classification system that is based on the SCOR model. He suggested that there exist 14 categories of risk, each of which are embedded within a particular part of the

SCOR model: Plan-Based Risks: Strategic, Inertia, Informational, Capacity, Demand; Source-Based Risks: Supply, Financial, Relational; Make-Based Risks: Operational, Disruption; Delivery-Based Risks: Customer; Return-Based Risks: Legal; Other-Risks: Environmental, Cultural (Rangel, de Oliveira, & Leite, 2015).

Despite multiple well-crafted attempts to standardize the classification, authors to this day still have yet to agree on a dominant classification system. For example, as recent as the year of the writing of this dissertation, yet another classification system was proposed by Khojasteh-Ghamari and Irohara. After conducting their own careful review of the literature, risks were categorized based on being supply-based, demand-based, or process-based risk (Khojasteh-Ghamari & Irohara, 2018). Time will only tell which, if any, classification system will be standardized and dominantly accepted by academics and practitioners alike. For this dissertation, I will not adopt any system of categorization for the sake of avoiding debate in an area that is not the focus of this work. Instead, I shall strictly focus on the category of "structural risk". This category can be mapped to any of the current classification systems proposed via careful consideration. Therefore, the contributions of this dissertation will not be lost with any generality of current or future classification systems.

Supply Chain Risk Identification

Identifying the type of risk that a supply network is exposed to is only one dimension of a larger framework of *risk identification*. Indeed, in Kleindorfer and Saad's SAM framework, not only should risks be categorized, but also identified within a network. Countless research has been conducted in this area of the literature. Earlier research in the area was from (Chapman, Christopher, Jüttner, Peck, & Wilding, 2002). They suggested that risk identification should follow a "3-P" approach (Philosophy, Principles, and Processes). They argue, from a philosophical perspective, that if firms seek to achieve supply chain continuity, they must embrace "...risk awareness for their top managers;... an understanding that changes in business strategy change supply-chain risk profiles;... that risk management is an integrated part of SCM;... and that each individual employee in each entity must have: a) risk awareness; b) an understanding

of his or her role in the risk management process" (Chapman et al., 2002, pg. 63). From a principles perspective, they argue that "risk considerations should influence the supply-chain design and structure,...,risk management should be based on a high level of supply-chain visibility and understanding amongst all entities, ... risk management should be based on clear performance requirements and lines of communication between all entities, ... [and] supply-chain risk management should be based on process alignment and co-operation within and between the entities" (Chapman et al., 2002, pg. 63). Last, from a process perspective, they argue that a collection of activities such as a formal risk identification process, assessment process, continuity and coordination management process, and learning processes, should be established if firms seek to address supply chain disruptions.

Juttner had later built on this framework by conducting a survey of practitioners to not only validate the model but also gain an understanding of the current state of SCRM from practitioners at the time (Jüttner, 2005). Generally, issues arise with the construction of a standardized risk identification framework. Gaudenzi had noted two such problems, one being that risk is present at various levels of the organization and the supply network environment, and secondly that risk evaluation is subjective, since "each analyst has his or her own concept of what constitutes a risk and of what is the nature of the upstream and downstream relationships" (Gaudenzi & Borghesi, 2006, pg. 114). Gaudenzi instead opted for a more pragmatic and quantitative approach by applying the Analytic Hierarchy Process (AHP) to risk identification and priority. Not only was this one (of many) methods to quantify a "risk", but also as an early attempt to remove the notion of subjectivity naturally within the inquiry "which risks are important".

This trend of "subjectivity-free" risk identification continued in the literature stream with more quantitative approaches. For example, Neiger suggested applying the value-focused process engineering methodology to supply chain risk identification. The general idea is to link value-focused decision modeling with extended-event-driven process chain modeling. He proposed five steps to achieve this process of identification, which involved risk source identification, risk measurement, and connecting this with various

objectives that a firm sets (Neiger, Rotaru, & Churilov, 2009).

Other researchers have suggested that risk identification is partly dependent on the perceptions of actors within the firm and supply network (G. A. Zsidisin & Wagner, 2010). In addition, other approaches have been suggested. A Bill-of-Materials approach was proposed as a way to determine risks from a manufacturing perspective (Takata & Yamanaka, 2013). So too was the approach for surveying a stake-holder network for outside perceptions of sustainability risk identification (Busse, Schleper, Weilenmann, & Wagner, 2017). Other more quantitative approaches have been suggested as well, such as extensions of the AHP approach (Song, Ming, & Liu, 2017) and the use of gathering and analyzing big data (K.-J. Wu et al., 2017; Sommerfeld, Teucke, & Freitag, 2018). In addition, Petri-Nets have been suggested as a possible way to quantify the risk identification process of more event-driven decision-support systems within the supply chain (Liu, Liu, & Liu, 2018).

Supply Chain Risk Assessment

Once risks have been identified, the next step is to assess them for their relevance and potential negative consequences to determine what, if any, mitigation strategies should be employed. Again, the current literature in this area is quite large and diverse. Early authors in the field had suggested employing methods from reliability theory such as fault-mode analysis, event-based analysis, decision-support analysis,interpretive structural modeling (ISM) , and AHP. (Gilchrist, 1993; T. Wu, Blackhurst, & Chidambaram, 2006; Nishat Faisal, Banwet, & Shankar, 2007). Others include Chance Constrained Programming (CCP), Data Envelopment Analysis (DEA), Multi-Objective Programming (MOP), and Petri-Net Modeling (D. Wu & Olson, 2008; Pfohl, Gallus, & Thomas, 2011; X. Wang, Chan, Yee, & Diaz-Rainey, 2012; Zegordi & Davarzani, 2012; Aqlan & Lam, 2015a, 2015b; Torres-Ruiz & Ravindran, 2018).

Other methods include the use of graph theoretic measures to quantify the vulnerability of a location in a supply network (Wagner & Neshat, 2010; Y. Kim et al., 2015). More recently, two trends have emerged suggesting newer methods to be used within risk assessment. The first is that of the Monte Carlo Simulation. A few studies have

already employed this method by simulating faults and disruptions within networks and quantifying the consequences of the effects (D. Wu & Olson, 2008; Klibi & Martel, 2012; Vilko & Hallikas, 2012; T. Wu, Huang, Blackhurst, Zhang, & Wang, 2013; Winter, Deniaud, Marmier, & Caillaud, 2018).

The second recent trend is the leveraging of Bayesian Networks to not only measure risk, but also risk-dependencies. This is a well-needed tool for the relatively new field of supply chain risk propagation assessment. Bayesian Networks, as I will explain in more detail later, model the interdependencies of random events via conditional probability distributions. The resulting network could then be used to construct risk measures, as well as conduct simulations to find vulnerabilities, and be used as a real-time decision support system for risk monitoring (and inferring the next most likely event to be influenced by a disruption) (Garvey et al., 2015). Some authors have already leveraged Bayesian Networks as a way to assess a variety of different risks as well as their interdependencies (Lockamy III & McCormack, 2010; Garvey et al., 2015; Qazi, Quigley, & Dickson, 2015).

Some Common Mitigation Strategies

Once the risks have been identified and assessed, the next step recommended by Kleindorfer is to design the proper mitigation strategy to address the risks. Again, this field of study in the extant literature is vast, and the methods range over strategic, tactical to operational. Just as with the taxonomy of risks, there seems to lack a consensus on a taxonomy of risk mitigation strategy. Some believe that risk mitigation should be seen from one of two primary approaches: preventative and reactionary.

Wagner and Bode suggested this approach to what he refers to as "Cause-Oriented" and "Effect-Oriented" SCRM Practices. Cause-Oriented strategies are more "top-down" approaches, where one concern their efforts at identifying and reducing the potential causes of a disruption. (Wagner & Bode, 2009; Nyoman Pujawan & Geraldin, 2009). The idea is simple from this perspective: firms should avoid activities that are "risky". A second approach is the Effect-Oriented strategy, where firms instead focus less on reducing the likelihood of an event and more on reducing the potential negative

consequences of the event.

An example of Cause-Oriented Mitigation (i.e. preventative) is that of mitigation of cultural risk via understanding the barriers of cultural understanding (Jia & Rutherford, 2010). On the other hand, an example of Effect-Oriented Mitigation is that of *buffering*, which is "an effort to reduce a firms exposure to the current exchange partner and to mitigate the detrimental consequences of disturbances that the relationship may confer" (Bode, Wagner, Petersen, & Ellram, 2011, pg. 836).

Other classification systems involve conducting the mitigation strategy based on the type of risk that is to be addressed. Since I have argued that there is no consensus within a supply chain risk classification system, there equally will be no consensus within a supply chain risk mitigation strategy classification system. As such, these other more complex systems of classification that rest on the type of "risk category" include a variety of strategies. For example, Chaghooshi identified over 41 different risk mitigation strategies in the extant literature (Chaghooshi, Momeni, Abdollahi, Safari, & Kamalabadi, 2018). Some examples include strategic stock, excess and reserved capacity, a flexible supply base, an economic supply incentive, multiple routes, postponement, distributed power, and many others. To discuss all of these at length would be outside the scope of this dissertation.

However, Kilubi had attempted to synthesize and categorize the variety of mitigation strategies within the extant literature. Upon an extensive analysis of the literature, he took some of the categories mentioned by Chaghooshi (albeit, independently), and assigned them to one of two categories: supply risk/demand risk. The strategies considered were: visibility and transparency, flexibility, relationships, postponement, multiple sourcing, flexible contracts, redundancy, collaboration, joint planning and coordination. The supply dimension was then crossed with the demand dimension, to create a 2 by 2 matrix and hence resulting in 4 general "supply chain risk mitigation" strategy categories. The "demand risk" was similar to others in regards to being proactive vs reactive (Kilubi, 2016).

While I will not diverge on a discussion of each and every possible mitigation strategy, I will point out that it is within the interest of this dissertation to discuss current

strategies in regards to supply network vulnerability, resiliency and robustness. That is, risk mitigation strategies that are centered around the design and redesign of the supply network via its structural properties.

A few of the articles in the extant literature have currently addressed these types of strategies. Behzadi tested the strategies of backup suppliers, moderating disruption probabilities, and the use of multiple suppliers within the network. It can be argued that these are all structural-based mitigation strategies (Behzadi, O'Sullivan, Olsen, Scrimgeour, & Zhang, 2017). Giaglis suggested that real-time vehicle routing is an effective transportation-based mitigation strategy (which would lend itself to a structural-based strategy in regards to mitigating "arc-risk") (Giaglis, Minis, Tatarakis, & Zeimpekis, 2004). Zhalechian designed a variety of risk-mitigation strategies when a network-design is of the hub-and-spoke form (Zhalechian, Torabi, & Mohammadi, 2018). Zokaee considered the problem of optimal network design under demand, supply, and transportation cost uncertainty (Zokaee, Jabbarzadeh, Fahimnia, & Sadjadi, 2017). There are many more types of structure-based risk mitigation strategies in the extant literate. I will direct the reader to Kainuma, who conducts a decent literature review on the matter (Kainuma, 2018).

2.1.3 Supply Chain Risk Measurement

Imperative to risk mitigation and assessment is the ability to properly measure and quantify supply chain risk. The notion of measurement of risk is still debated in the extant literature, as it is still in it's infancy. Yet again, no consensus seems to have been reached on this topic. This may be due to the ever-changing nature of SCRM. As such, I will review through some of the prior approaches to measuring supply chain risk. I will then argue that the measures proposed by Garvey, Carnovale and Yeniyurt are the most appropriate for supply chain risk management, as their measurement framework is general enough to apply for every type of risk that the supply chain risk manager is likely to encounter, regardless of level or experience. In addition, their approach allows for the combination of subjective likelihood estimation by managers, as well as objective measurements recorded by monitoring technology.

Prior Attempts to Quantify Supply Chain Risk

One of the more popular methods to measure and analyze supply chain risk is the Analytic Hierarchy Process (AHP). This method is inherently a multi-criteria decision-making methodology, but has been applied to the measurement of supply chain risks in recent years. Among the first was (T. Wu, Blackhurst, & Ogrady, 2007). The method proposed was to first categorize risk based on a manager's perceived level of severeness. The likelihoods of the risks occurring were also captured by expert opinion. These two concepts were taken into account using the traditional formula of risk: Likelihood $\times$ Severity $\times$ Detection. The AHP method then takes this information and organizes it into a matrix by looking at pair-wise comparisons. The analyst then calculates a priority vector, defined as the Eigenvector of the largest Eigenvalue of the matrix. The priority vector is a ranking of the risks, which contains "measures of risk" for each factor or source of the event.

Extensions of the AHP model for risk measurement have also been proposed over the years. The notion of "uncertainty" of the expert opinion was controlled for by leveraging fuzzy set theory in combination with an AHP approach (X. Wang et al., 2012; Samvedi, Jain, & Chan, 2013; PrasannaVenkatesan & Goh, 2016). It was also extended by combining it with a failure mode and effects model (P.-S. Chen & Wu, 2013). In addition, the AHP method had been applied within industry-specific risk source measurement as well as multi-dimensional consequence impact measurement (Q. Dong & Cooper, 2016; PrasannaVenkatesan & Goh, 2016; Mital, Del Giudice, & Papa, 2018).

Yet another tool that has been proposed for risk measurement is the Data Envelopment Analysis (DEA). This approach differs from AHP in that efficiency scores are assigned to decision making units (DMU) and then a sequence of linear programs are solved to identify the "most efficient" unit. From there, other units are compared to this most efficient unit, and inefficient units are able to be identified. The approach has been applied to various decision making problems within supply chain (D. Wu & Olson, 2008; T. Wu, Shunk, Blackhurst, & Appalla, 2007; Weng, Wu, Blackhurst, &

Mackulak, 2009). In addition, it has been applied to supply chain risk measurement approaches and decision making problems (Talluri, Narasimhan, & Nair, 2006). For application to risk measurement, typically a random variable is introduced into the set of constraints for the primary DEA model.

An approach that attempts to capture inter-dependencies of risk, is the Interpretive Structure Model (ISM). This approach gathers information from various risk sources and creates a dependency graph. The graph is then converted into a matrix, where then it is converted into an equation. Each instantiation of the risk is then replaced into the equation, where a "risk score" is calculated for each respective risk. This approach has been applied to identifying sources of information risk (Nishat Faisal et al., 2007) as well as more general categories of risk (Pfohl et al., 2011; Nguyen et al., 2018). Similar graph-theoretic approaches have been proposed that is analogous, yet more general, than ISM (Wagner & Neshat, 2012; Nakatani et al., 2018).

Zegordi argues that "the main assumption [of current risk measurement] is risks patterns, can be estimated as different distribution functions, these estimations are based on previous risk experiences and existence of definable behavior" (Zegordi & Davarzani, 2012, pg. 2104). As such, he argues that disruption analysis (and therefore risk analysis) should be conducted within a deterministic framework. Such a framework is the Petri-Net model, proposed by Carl Petri. This approach is more of a system-engineering and reliability modeling approach. The idea is that system (in this case, the supply chain) is modeled with different types of nodes (called place and transition nodes), and different events are "fired" throughout the system. The results of the events can then be analyzed for reliability and efficiency purposes. This method was applied by Zegordi, among others, to model supply chain disruptions and the anticipated subsequent consequences (Zegordi & Davarzani, 2012; Zhang & Lam, 2016; Blackhurst et al., 2018; Liu et al., 2018).

A variety of other approaches have also been suggested. The idea of these other approaches rest on an "expert level" of knowledge of the risks themselves. The idea is that a group of experts are surveyed to estimate the likelihood of an adverse event and the subsequent potential damage of the event, where "damage" is measured a variety

of different ways (estimated cost, estimated number of nodes, estimated environmental damage, scaled response, etc). The common approach to these methods is to create an index for both likelihood and severity, respectively, via survey methods. The indicies are then used to calculate composite risk measures for each source of risk. One such example of designing this composite score is to simply multiply the severity index score by the likelihood index score. Other methods involve just a simple scaled response for the construct "risk"(Tummala & Schoenherr, 2011; Baharmand, Comes, & Lauras, 2017; Kumar, Basu, & Avittathur, 2018; Vilko & Hallikas, 2012; Varzandeh, Farahbod, & Zhu, 2016).

More in line with the mathematical notion of "risk" are methods that quantify it with probability distributions. A "risk" is simply modeled as a random variable, which of course has a collection of outcomes and an associated probability distribution. The modeling process varies. A modeler could designate the random variable to represent whether or not a particular event occurred (thus holding a value of 1 or 0), and assign a simple Bernoulli distribution to the variable. Separately, the consequence of the risk can be a cost function of the variable itself. Another approach is to model the consequence of the risk itself using the random variable. That is, the outcome of the consequence can lie on a numerical continuum (or within a discrete collection of potential losses), and the probabilities of those losses are described via the distribution of the random variable. A third approach, common in financial risk modeling, is to model a loss/gain of a "risk" as a random variable, where both positive values and negative values are possible outcomes for the random variable. These modeling approaches have been the standard approach in the process of incorporating the notion of "uncertainty" into a variety of operations research and supply chain models (Klibi & Martel, 2012; Han & Shin, 2016; X. Chen, Xi, & Jing, 2017; Wagner, Mizgier, & Papageorgiou, 2017).

While the random variable approach to modeling both the consequence and likelihood of a supply chain risk is more in line with it's mathematical definition, there are many other tools and measures of risk that are derived from the random variable itself. Simply modeling the risk as a random variable is not enough. As any student of probability knows, the random variable is just that, random. In order to gain understanding

of the underlying phenomena that the RV is representing, well-defined measures need to be designed from the RV to make sense of it. One approach, and the most common, is to find the expectation of the random variable (Snyder, Scaparra, Daskin, & Church, 2006; Nooraie & Parast, 2015; Meena & Sarmah, 2016; Nooraie & Parast, 2016; Govindan & Fattahi, 2017). Such an approach allows for a deterministic sense of the underlying phenomenon, as well as it's ease of incorporation into decision making modeling such as mathematical programming approaches. Two other possibilities are to look at the "risk", derived from the random variable, from a Value-At-Risk (VaR) perspective (Mehrjoo & Pasek, 2016; Mizgier, Pasia, & Talluri, 2017). In addition, the RV can be used to estimate a "worst-cost" scenario (Snyder et al., 2006).

Another application of the random variable approach is to consider decision problems of risk measurement within the scope of what is known as the Real Options Approach. This approach is derived from the concept of a financial option. For example, if a perspective buyer of a stock would like to purchase said stock, but feels that the "risk" of doing so at the moment is too high, he can instead opt to purchase a contract, rather than the stock itself. The contract value is itself significantly lower than the value of the stock.

For example, 100 shares of Apple may cost him now $19,000. If he purchases the stock, and the value of Apple decreases to $150 per share, this would equate to a loss of $4,000. Instead, the buyer can buy a contract, that gives him the right, but not the obligation, to purchase the 100 shares of Apple. The contract would cost significantly less. For example, for a $195 contract that expires in four weeks, this would probably cost him a total of $200. The "risk" of the buyer purchasing Apple stock is significantly higher than that of the contract. If the stock indeed drops to $150 per share, the buyer only sees a loss of his full $200, rather than the $4,000. If the stock goes higher to $220 per share, the buyer can purchase the stock at $195 per share (due to the contract) and sell it for $220 (for a profit of $25 per share, or, $2500 - $200 = $2300).

The same approach can be used for measuring risk within supply networks for certain types of mitigation strategies. For example, Carbonara used the real options approach to analyze the strategy of postponement (Carbonara & Pellegrino, 2017). The idea

is simple: a manufacturer could decide to design their process in a way that can be changed to produce a different product. The investment into this level of flexibility is analogous to the "contract cost" in the options framework. During a disruption, the manufacturer has the "right", but not the "obligation" to adjust the process during a disruption to produce a different product. The "exercise" of the contract is the decision to shift manufacturing to the new product. The "expiration" of the contract is the decision to continue current manufacturing of the current product. The "strike price" and "underlying price" are analogous to the current state of the supply chain (that is, whether or not a disruption has occurred). This interpretation of the process allows the analyst to incorporate more of a financial risk perspective to supply chain risk measurement. That is, the analyst can determine the measurement of a supply chain risk by employing familiar types of financial risk management in options theory (such as optimal time to exercise, "strike price" to buy, optimal number of contracts, etc).

A more recent approach to supply chain risk measurement is the use of Bow-Tie Analysis. This approach fuses together the use of random variables and that of fault-tree and event-based modeling approaches. The mechanism behind this analysis is to map the "risks" into nodes of a network. The dependencies are then drawn. Some nodes are "risk enhancers" while others can be modeled as "risk reducers". These essentially can represent management's various mitigation (or other decisions) strategies. The risks themselves are modeled as random variables that have probabilities of occurring. It is often assumed that these probabilities must be independent of the other risks. Risk scores for particular actions and events can then be analyzed by taking into account these dependencies and random variables. While the approach has been traditionally applied in studying operational risk, it has recently been employed to measure supply chain risk (Aqlan & Lam, 2015b, 2016).

Last, and more recently, the use of Bayesian Networks have been applied to supply networks as a means to an end: to construct meaningful risk measures. The dynamics of Bayesian Networks will be discussed later. The idea is simple however. Risks are modeled as random variables. They can be discrete or continuous. Each risk is also

modeled as a node in a directed graph. The "structure" of the network is determined by drawing directed arcs between the nodes. These determinations are made either by expert opinion, through learning algorithms from data, or a combination of both. Once the graph is drawn, conditional probability distributions are determined, again either by expert opinion or via learning from data.

The resulting structure is called a *Bayesian Network*. This network can then be used to conduct inquiries. If we know that a particular risk has manifested, we then call the manifestation *evidence*. Upon the existence of evidence (either through one or more risks), we can *update* our beliefs, or probabilities, of the other events in the network. The resulting structure can hence be used for risk measurement calculation, disruption simulation, real-time risk monitoring or as a knowledge-based tool of risk. That is, it is very flexible and can easily apply to many different areas of study.

Bayesian Network's applications to supply chain risk measurement is nothing new. Pai was the first to propose their use (Pai, Kallepalli, Caudill, & Zhou, 2003). Unfortunately, the literature did not seem to adopt their methodology, nor did it's potential usefulness for modeling the interdependencies between risks be seen be academics. Soon thereafter, Lockamy had applied Bayesian Networks to SCRM by leveraging them to create "risk profiles" for firms. The idea was to take information regarding different risks inherent within each firm and use a Bayesian Network to create risk profiles. The risk profiles were then used in tandem with a Value at Risk approach to calculate a VaR for each firm (Lockamy III, 2011; Lockamy III & McCormack, 2010, 2012).

Recent expansion of this area of research has been expanded by two authors: Garvey and Qazi, who independently decided to leverage Bayesian Networks for the goal of designing risk measures of propagation and resiliency. (Garvey et al., 2015) proposed an approach that involved (1) taking each location in the supply network (both node and arc) and identifying the collection of all possible risks in those locations, (2) identifying the interdependencies of the risks within each location and across the locations, (3) use data or expert opinion to estimate the conditional probabilities of the risks identified, and last (4) use the resulting Bayesian Network (which they coined as the Risk Graph) to construct risk measures.

The measures they proposed are inherently measures of risk propagation. Their model is general enough to incorporate both causal risk and diagnostic risk measurement, as well as have the ability to construct other types of risk, resilience and robustness measures. The model is flexible enough to incorporate subjective and objective knowledge, as well as apply to a wide array of applications, at a variety of levels of analysis, for a variety of areas within SCRM. In other words, their model is the most flexible in the extant literature to be applied to almost any supply chain risk problem that requires measurement of supply chain risk, and it could be argued that all other models of measurement proposed thus far can be represented within the context of their model (Garvey et al., 2015).

Independent of (Garvey et al., 2015), (Qazi, Quigley, & Dickson, 2015; Qazi, Dickson, Quigley, & Gaudenzi, 2018; Qazi, Quigley, Dickson, & Gaudenzi, 2015; Qazi, Quigley, Dickson, Gaudenzi, & Ekici, 2015; Qazi, Quigley, Dickson, Gaudenzi, & Önsel, 2015; Qazi, Quigley, & Dickson, 2018; Qazi, Quigley, Dickson, & Ekici, 2017) also applied Bayesian Networks supply chain risk measurement and propagation . They extended (Garvey et al., 2015) framework by proposing new risk measures that are dependent on the risk factor type itself. They argued that (Garvey et al., 2015) model had a major shortfall. Indeed, they state that the "association of this measure (Garvey et al., 2015) with each element of the supply network rather than risk factor itself results in a major problem because for all different combinations of instantiation of risk factors corresponding to each element, propagation across the pure descendants only depends on the state of specific parent node that connects risks of that element to the rest of risk factors" (Qazi, Quigley, Dickson, Gaudenzi, & Önsel, 2015, pg. 426). However, I will argue later that such criticisms will not threaten the strength of this research.

2.1.4 Supply Chain Risk Propagation (SCRP)

Risk Propagation vs. Disruption Propagation

A recent trend in the literature has been the emergence of the construct *risk propagation*. Unlike traditional risk management, which views risk as the potential to directly

manifest into a disruption, risk propagation seeks to understand the sequence of events that lead to the disruption itself (with the "events" not necessarily being disruptions) as well as the potential to cause other disruptions (Garvey et al., 2015). This recent trend is important to understand from a managerial viewpoint, as it would allow for the proper mitigation of risks that lead to other risks of which eventually "trickle down" to the primary driver of the disruption.

There is some confusion and intertwining of the notions of "disruption" and "risk" in the extant literature. Recent literature has attempted to disentangle these two constructs. (Garvey et al., 2015) began by suggesting that a supply chain risk is an adverse event that can manifest into a disruption (Garvey et al., 2015). This is further advocated by (DuHadway et al., 2017), who offers a classification of disruption itself.

Disruption propagation focuses on the sequence of events that occur due to a disruption itself. That is, it's focus is on what the disruption itself portends. If a system such as a logistics network or a computer network is of a structure that was intentionally designed, then the designer has intentionally placed dependencies within the system itself. Hence, if a disruption were to occur, then the disruption should be expected to spread throughout the rest of the designed system. Understanding the planned inner-connections between these various components of the system is of no doubt important. Hence, a disruption propagation analysis seeks to understand the deterministic failures in the event of a single failure in the system (this is, at it's core, the reason for reliability theory). This form of analysis makes itself an appropriate candidate for applications of reliability-theory-based and event-driven-based analytical frameworks. The deterministic nature of the various management activities and physical locations allow for the ease of use for these types of analytical frameworks, as has been demonstrated in the prior literature (Wagner & Neshat, 2012; Nakatani et al., 2018; Zegordi & Davarzani, 2012; Zhang & Lam, 2016; Blackhurst et al., 2018; Liu et al., 2018).

The reason as to why prior literature has applied deterministic-based analytical frameworks to supply-network risk propagation appears to be due to a lack of proper distinction between risk propagation and disruption propagation. I will clarify them now more formally with a definition that I feel is most appropriate given the nature

of supply networks. *Supply chain disruption propagation* is a sequence of deterministic events that *will* occur in the event of an originating disruption due to the design of the supply network. Here, "design" encompasses not only structure, but also mitigation policy, strategic decisions, best practices, people, and processes. That is, more generally, anything within the network that was knowingly and purposely implemented or integrated.

On the other hand, *supply chain risk propagation* is the sequence of events that *could* occur, which may (or may not), lead to a disruption propagation or a different risk propagation. Hence, supply chain risk propagation encompasses supply chain disruption propagation, but is more general and encompassing of the environment of which is external to not only the firm, but also the supply network itself. There are two points of emphasis I would like to make in this definition. First, the use of the word "event" is to be understood as either a risk or a disruption. Second, I use sequence in the most general sense possible, in so being that a sequence of "disruptions" followed by "no disruption" does not necessarily imply that subsequent nodes will therefore have "no disruption".

A Review of Risk Propagation

Risk propagation has gained attention over the past few years, as authors have began to recognize that it's impacts and forms of analysis diverge from more traditional risk assessment methods. In the era of big data and blockchain, firms have an ever-growing quantity of data available to them regarding the structure and various operations within their supply networks. This observation is personified by the current literature on risk propagation. One of the early works in the literature is from Ghadge , who constructed a conceptual framework of risk propagation (Ghadge, Dani, & Kalawsky, 2011).

His dissertation work later presented an early analytical approach to analyzing this form of risk and disruption, as well as explain some of the difficulties in modeling such an issue. His approach was to use data to estimate the probabilities of various risks within the network. He subsequently ran a simulation and computed final risk measures based on probabilities of cost and time impacts (Ghadge, Dani, Chester, & Kalawsky,

2013). In addition, early attempts to model risk propagation had their roots in Shin (Shin, Shin, Kwon, & Kang, 2012). They developed a model that is similar to (Garvey et al., 2015), where they consider key risk indicators in each location of a supply network leveraging Bayesian Networks.

After this time, the risk propagation literature seemed to attract interest from academics across multiple different methodological perspectives. Graph-theoretic approaches were proposed by leveraging network characteristics in the use of risk propagation and robustness measures (Olivares Aguila & ElMaraghy, 2018; Bakshi & Mohan, 2015). Structural equation modeling and statistical approaches were offered as yet another means to measure the affects of propagation (Truong Quang & Hara, 2018; Ghadge et al., 2013).

Simulations also began to attract interest in the literature. Ivanov proposed using a systems approach to model the supply network and to randomly generate disruptions and random recovery efforts. He measured a variety of resulting key supply chain performance metrics to determine the various impacts of a distinct, yet related, phenomenon known as the "ripple effect" (Ivanov, 2017). Others conducted simulations of risk communication efforts (Jiang, Liu, Huang, & Zhang, 2017), impact of trust and topology (Hou, Wang, Wu, & He, 2018) and mitigation efforts of propagation (Z. Wang, Hu, & Gong, 2018).

Others in the literature took more of a traditional risk management approach and applied it to propagation. Value at risk was used to measure the propagation impacts in the case of information sharing (Xu et al., 2015) and being used in sensitivity and risk aggregation (Mizgier, 2017). In addition, some have leveraged the concept of Shannon's Entropy, which also was used to quantify the impacts of the ripple effect (Levner & Ptuskin, 2018).

However, despite these prior attempts, two primary methods of analysis grew within this stream of literature that currently seem to be dominant. Given the "cause-and-effect" nature of disruptions and the construct of propagation, academics from the complex-systems and decision-analysis literature seem to have applied traditional reliability and event-based modeling to measuring impacts of propagation. Govindan

proposed a multi-criteria decision making model to better understand the impacts of propagation (Govindan & Chaudhuri, 2016). The systems-based approach was also applied to risk propagation in quantification of the ripple effect (Sokolov, Ivanov, Dolgui, & Pavlov, 2016; Dolgui, Ivanov, & Sokolov, 2018). The same modeling approach was also applied to studying risk propagation's effect on reliability and resiliency (X. Chen et al., 2017; Valenzuela, Fu, Xiao, & Goh, 2018). In addition, the construct of "cascading failure" was studied within the context of systems-based models (L. Tang, Jing, He, & Stanley, 2016; Y. Wang & Zhang, 2018). There are many others in this literature stream as well that I have chosen to refer the reader to various literature reviews on the matter (Scheibe & Blackhurst, 2018; Brintrup & Ledwoch, 2018; Dolgui et al., 2018).

In addition to the reliability systems engineering approach to risk propagation, a second approach has emerged as a dominant tool. The application of Bayesian Networks seemed to have been proposed to fill a void that exists in the systems-reliability literature: incorporating uncertainty. As mentioned, Shin was among the first in the literature to suggest that Bayesian Networks can be used to model the interdependencies of risks within a supply network (Shin et al., 2012). (Garvey et al., 2015) synthesized the risk measurement literature to propose not only new supply chain risk measures, but a general framework that leverages Bayesian Networks to measure supply chain risk propagation (Garvey et al., 2015). Qazi, in a series of papers, had extended (Garvey et al., 2015) framework to address a variety of issues in risk management through the lens of risk propagation leveraging the framework proposed by (Garvey et al., 2015) (Qazi, Quigley, & Dickson, 2015; Qazi, Dickson, et al., 2018; Qazi, Quigley, Dickson, & Gaudenzi, 2015; Qazi, Quigley, Dickson, Gaudenzi, & Ekici, 2015; Qazi, Quigley, Dickson, Gaudenzi, & Önsel, 2015; Qazi, Quigley, & Dickson, 2018; Qazi et al., 2017).

Others had applied the Bayesian Network approach to measuring resiliency and robustness in supply networks while taking into account risk propagation (Hosseini & Barker, 2016; Han & Shin, 2016). In addition, others created extensions and other tangential ideas to the framework proposed by Garvey and later extended by Qazi (Ojha, Ghadge, Tiwari, & Bititci, 2018; Fu, Zhang, Montewka, Yan, & Zio, 2016; Daultani, Goswami, Vaidya, & Kumar, 2017). Hence, the risk propagation literature

is dominant in two primary methods of analysis: systems-based reliability analysis and Bayesian-Network-based risk analysis.

2.2 Fundamentals of Bayesian Networks

2.2.1 Conditional Probability and Independence

In basic probability theory, the notion of a *random variable* X typically represents a numerical value that could possibly be assigned to X at some point in the future, once observed. Every random variable has two primary components. Namely, the sample space S_X and a corresponding probability distribution $f(x)$ that describes the frequency of observations of the possible values that X can be assigned. A sub-branch of Probability Theory is that of the Bayesian Philosophy, Bayesian Statistics, or Bayesian Probability. All rest on the general principle that the idea of *probability* is inherently a *belief* that one holds, measured on a 0 to 1 scale (0 being less likely, 1 being more likely). The idea is that every individual has a collection of observations and concepts often referred to as *knowledge*. This knowledge, often represented as ξ, represents all the beliefs that one has about the world surrounding them, and is often referred to as *a priori* knowledge, or prior belief.

Bayesian Probability/Philosophy/Statistics is based on this fundamental view of probability. In addition, it is based on one more fundamental concept, that of *conditional probability*. Conditional probability transcends the concept of *belief updating*. For example, if X represents the amount of income we will earn next year, then an individual initially holds a belief that X will be a with a value of belief of $P(X = a \mid \xi)$. When an individual observes a new event e, say, such as a new boss entering the ranks, this belief is then updated to an *a posteri* belief, namely $P(X = a \mid \xi, e)$. To compute this, the standard definition of conditional probability is often employed:

$$P(X = a \mid \xi, e) = \frac{P(X = a, \xi, e)}{P(\xi, e)}$$

Typically, Bayes Theorem is employed in the denominator in order to obtain a

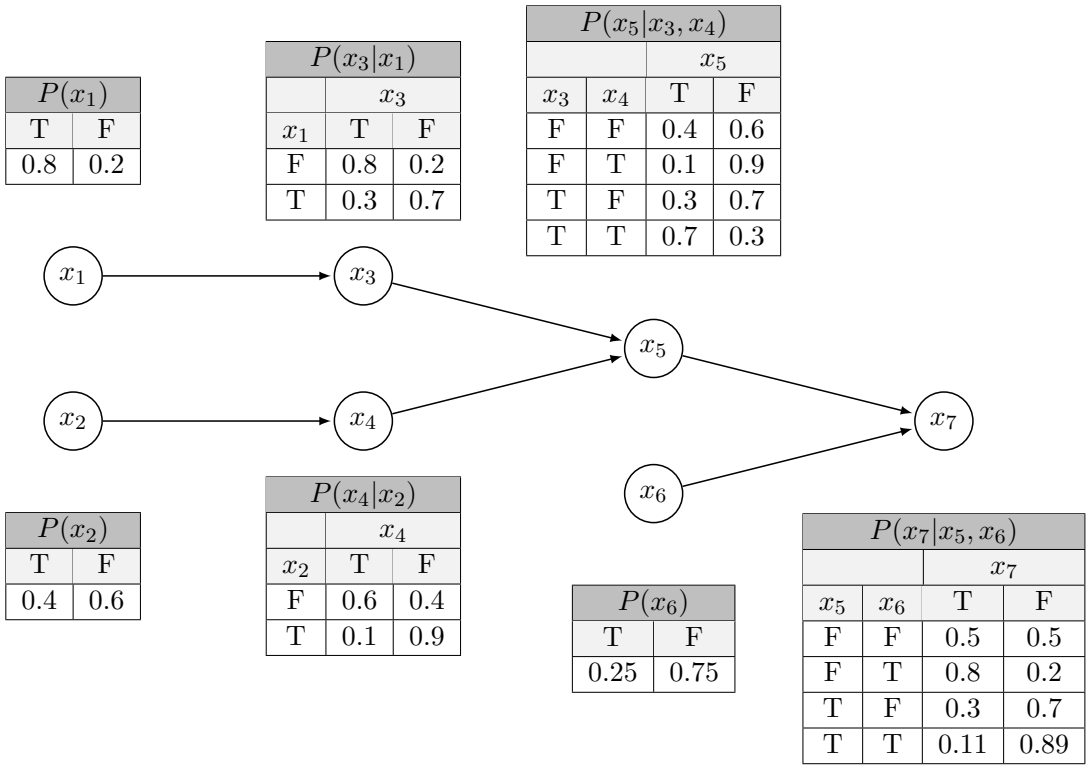
meaningful result. Namely:

$$P(X = a \mid \xi, e) = \frac{P(X = a, \xi, e)}{P(\xi, e)} = \frac{P(X = a, \xi, e)}{\sum_{i \in S_X} P(\xi, e | X = i) P(X = i)}$$

In probability, random variables can be understood to be *independent* if they do not exhibit a *dependence* on each other. Mathematically, if we have variables X and Y, then they are independent if $P(X|Y) = P(X)$. That is, if the belief that X will occur given knowledge of the variable Y is the same as having no knowledge of Y, then intuitively, the two variables are *independent*. In addition, two variables are said to be *conditionally independent* with respect to a variable C if they happen to both be dependent on C (and hence, they themselves may not be independent), but become independent when knowledge of C is known. For example, if A depends on C, and B depends on C, then we say that A and B are *conditionally independent* with respect to C if $P(A|B, C) = P(A|C)$. In other words, the knowledge of B in that instance is excessive and useless knowledge. One must of course be careful with this interpretation. If knowledge of C is unknown, it is still plausible for A and B to be dependent on each other. This is why we refer to this instance as *conditional independence*. That is, we only know that the two variables are independent when given information about another common variable.

2.2.2 Bayesian Networks: A Model of Belief Propagation

The model of dependency between random variables that I have illustrated above is quite simple to understand when the number of variables is very low. On the other hand, if we happen to have a large number of variables, then merely describing the dependencies descriptively or mathematically becomes a burden. This was the primary motivation behind the invention of the Bayesian Network. The Bayesian Network is a *directed acyclic graph*, where the nodes represent the random variables under consideration and the directed arcs represent probability dependencies between the variables. We must think of a Bayesian Network as essentially a *cause and effect diagram*. This is one reason why the resulting network must be acyclic. If we had a directed cycle, the notion of endogeneity would enter into our model of dependency. That is, we essentially



$P(x_1)$	
T	F
0.8	0.2

$P(x_3\|x_1)$		
	x_3	
x_1	T	F
F	0.8	0.2
T	0.3	0.7

$P(x_5\|x_3, x_4)$			
		x_5	
x_3	x_4	T	F
F	F	0.4	0.6
F	T	0.1	0.9
T	F	0.3	0.7
T	T	0.7	0.3

$P(x_2)$	
T	F
0.4	0.6

$P(x_4\|x_2)$		
	x_4	
x_2	T	F
F	0.6	0.4
T	0.1	0.9

$P(x_6)$	
T	F
0.25	0.75

$P(x_7\|x_5, x_6)$			
		x_7	
x_5	x_6	T	F
F	F	0.5	0.5
F	T	0.8	0.2
T	F	0.3	0.7
T	T	0.11	0.89

Figure 2.1: An example of a Bayesian Network (Garvey, Carnovale, & Yeniyurt, 2015).

would have a chicken and egg problem (that is, which came first?).

Once the random variables have been decided and the causal paths have been drawn between them (or, more appropriately, between the nodes that represent the random variable), then we must specify the conditional probability distributions for each variable. For example, in Figure 2.1 the variable x_7 is dependent on the variables x_6 and x_5. Therefore, in order to model the probabilities of the possible values for the random variable x_7, we must specify four distinct (though not necessarily unique) probability distributions for x_7, each one corresponding to a specific outcome of the combination of the values for x_6 and x_5. If these variables are binary random variables, then the only two values they can hold are 0 or 1 (or, F or T).

Therefore, there are four scenarios that are possible: $x_6 = 0, x_5 = 0$, $x_6 = 1, x_5 = 0$, $x_6 = 0, x_5 = 1$, $x_6 = 1, x_5 = 1$. Hence, we must define what the probability of x_7 would be in each of these four instances (in order words, we need to specify four probability distributions for the variable x_7, as well as all other variables in the model, in order for

it to be "fully specified"). If a node has n parents, then the number of distributions we must assign to it is 2^n (assuming all the variables are binary). The natural question that comes to mind with this specification of a model is, where do the probabilities come from?

2.2.3 Learning Bayesian Networks

Bayesian Networks are not just pulled out of thin air. Analysts do not just assign arbitrary probabilities for the 2^n requirements. These probability distributions come from one of two sources, which is what precisely makes Bayesian Networks very flexible to many scenarios: (1) We query "experts" in the field, ask them what they believe causes a certain variable, or, ask them to identify a cause and effect diagram based on their expertise. From there, a survey design can be implemented and we can use a scaled response survey to have these "experts" "guesstimate" the 2^n probabilities. (2) We can use a data set and estimate the probabilities using statistical estimation methods (Gibb's Sampling, MCMC, GMM, etc).

The beauty of Bayesian Networks is that since the random variables are assumed binary (they don't need to be, but for this dissertation, they are considered as such), each distribution can be itself modeled as a *Beta* distribution. So, once the structure of the Bayesian Network is known, and once the probability distributions of the the network is also known, we can represent the Bernoulli Distribution by converting it to a Beta distribution. The distinct advantage of doing this is having the ability to leverage *both* expert data and objective observational data (or both!).

Ideally, a system can be "calibrated" using expert opinion. Once the system "starts", it *learns* by observing new data and reestimating the parameters of the Beta Distribution. New probabilities are then stored. That is, the Bayesian Network can act as a *knowledge network* that is always updating it's beliefs given new information about the world. The source of the information is of no concern to the Network. Expert, objective, pulled out of the sky. It doesn't matter, the Bayesian Network can handle it all.

We can design the network structure in any way we wish. Typically, however, a

standard procedure is taken if we are attempting to find the structure of a Bayesian Network from a given data set. The first task is often referred to as *structure learning*. The second take is often referred to as *parameter learning*. Structure learning involves using the data and the defined variables (the nodes) to find an optimal structure (that is, the "best way" to combine the nodes with directed arcs), where the only constraint placed on the procedure is that it must render an acyclic graph. After the structure is known to the analyst, the analyst then takes the data and finds the *parameters* of the Bayesian Network. This could either be trying to estimate the alpha and beta parameters of the Beta Distribution, or, simply finding the conditional probabilities directly from the data using a simple proportion (since the simple proportion in this instance is the MLE estimate).

Typically, on a more mechanical level, the minimum descriptive length (MDL) and the Bayesian Information Criterion (BIC) are used to judge how good of a "fit" the structure is to the data. There are many algorithms that accomplish the estimation of network structure and the estimation of the network parameters (Heckerman, Geiger, & Chickering, 1995; Heckerman, 1998; Pearl, 2014; Garvey et al., 2015). I will avoid a discussion entirely of these since they are outside the scope of this dissertation, and I direct the reader to the aforecited references. The important idea to take from this review is that Bayesian Networks represent random variables, of which have a dependency structure on them, that we are seeking to leverage in order to conduct something known as *probabilistic inference* (Pearl, 2014).

2.2.4 Belief Inference and Propagation

Once we have a fully specified Bayesian Network, the next natural question is to ask, how shall we leverage it? One of the fundamental philosophical reasons for constructing the Bayesian Network is that it was intended to act as means to reason *probabilistically*. That is, to make judgments about certain events or outcomes of scenarios when certain information is provided to the user. Pearl refers to this process of attempting to understand the "likelihood", or, more apropos, the "belief" of an event or scenario as *probabilistic reasoning* (Pearl, 2014).

Bayesian Philosophy rests on the foundational philosophical axiom that all humans are inherited with a certain innate set of knowledge. The brain essentially holds a memory of this knowledge by remembering a combination of *concepts* (which in our application, would the be nodes in the Bayesian Network of which represents the probabilistic events), or relations between concepts (that is, the notion of logical inference).

For example, if we happen to see that many times when it is cloudy out, rain may follow soon after. We may remember this, that is, *learn* this by accomplishing a few mental tasks. First, what does it mean to be "cloudy"? What does it mean to be "raining"? Once our brains have a conceptualization of the concept at hand (what Plato referred to as *ideas*), we then often conduct *inductive logic*. That is, our brains have a tendency to generalize repeated occasions where two concepts happen to occur close to each other in time (Russell, 1919), and hence we remember these observations, and therefore have the ability to conduct *inductive inference* (Russell, 1912).

Hence, after we have "learned" the general attributes of "sky", "clouds", "rain", our minds organize these concepts via a temporal ordering, where it remembers what will subsequently follow when one of the concepts are realized. If we realize a concept, we then *infer* that the related concept will occur (Occam's Razor suggests that our minds will think that if something is *likely*, then it is *certain*, despite the obvious problem with such a conclusion. Rather, as Popper argues, we just simply need to consider it as a proportion of times something has occurred (Popper, 1959). Nothing more, nothing less).

For example, after a sufficient number of observations of the two simple concepts of "clouds" and "rain", we tend to believe that it will rain if we happen to see clouds. If one has woken up during the beginning of the day, one may expect the probability that it will rain during the day to be quite low (say 0.20). The moment they step outside and look up to see dark clouds, their *belief* of it raining has changed, and the chances of it has increased. That is, the individual has *reasoned probabilistically* that upon the observation of *evidence*, namely the clouds, they must take into account this evidence and update their probabilities.

These are often referred to as *subjective probabilities*, since the likelihoods of the

event occurring are partly dependent on prior knowledge and recent evidence observation. Philosophically speaking, the *Objectivist* would of course disagree with the general idea that a probability is somehow *subjective*. The *Bayesianist* truly believes that their observations have been correct, and hence, upon the sight of new evidence, they can conduct *inference* using that evidence and their own set of prior knowledge.

To illustrate the process of *belief updating*, let me formalize this more mathematically. First, let C denote the event that it is cloudy (0 if not, 1 if it is). Let R represent the event that it is raining (0 if not, 1 if it is). Suppose that based on prior knowledge and your own personal observations, you estimate that the probability that it will rain, given that there are no clouds, is 0.1. One the other hand, assume you estimate the probability that it will rain, given that there are clouds, is 0.6.

With Bayesian Networks, if the the events are modeled using binary states (happens, does not happen),relative to the observer, there are really three states: happens, does not happen, no knowledge. For example, if you happen to be inside the whole day and would like to know the probability it will rain, then you have no evidence to use to calculate $P(R)$, but rather, only prior knowledge. What you do have, however, are the conditional probabilities. Thus, using the law of conditional expansion, we can write $P(R = 1) = P(R = 1|C = 0)P(C = 0) + P(R = 1|C = 1)P(C = 1)$. Suppose you *believe* that the probability it will be cloudy outside is 0.4. This is called your *belief*, also known as an *a priori* belief (or more mechanically, an apriori distribution). If this is the case, then you know that $P(C = 0) = 0.6$, $P(C = 1) = 0.4$. Hence, you should be able to easily compute $P(R = 1)$:

$$\begin{aligned} P(R = 1) &= P(R = 1|C = 0)P(C = 0) + P(R = 1|C = 1)P(C = 1) \\ &= (0.1)(.6) + (.6)(.4) \\ &= 0.3 \end{aligned}$$

On the other hand, suppose you happen to know that it is indeed cloudy outside.

Then, you would have the following:

$$
\begin{aligned}
P(C=1) &= 1 \\
P(C=0) &= 0 \\
P(R=1) &= P(R=1|C=0)P(C=0) + P(R=1|C=1)P(C=1) \\
&= (0.1)(0) + (.6)(1) \\
&= 0.6
\end{aligned}
$$

That is, we obtain two different probabilities for the same exact event. So what was different between the two? Namely, the observation of evidence available to us. In the first example, we did not have exact knowledge regarding the clouds, and instead speculated that $P(C=1) = 0.4$. That is, we had a *prior belief* in what we thought the probability of it being cloudy outside was simply because we didn't have any evidence or knowledge on the event regarding the clouds. On the other hand, in the second case, we have complete and exact knowledge. We may have walked outside, observed it was cloudy, and shifted our belief of it being cloudy from 0.4 to 1.0.

This form of *probabilistic inference* is often what is called *causal analysis*, or, *probabilistic induction*. We can think of this as *forward* inference, since we start with beliefs at "the top" and "work down" the chain of causality. However, the advantage of Bayesian Networks is that we can also work *backwards*, through a process known as *abduction*. The classic example of this entails the character of Sherlock Holmes, of whom is famously known for his "deductive ability".

This is actually a misnomer, as no where in the novels does it portray Holmes of having *deductive ability* (other than by the use of the term), but rather, *abductive ability*. Deductive ability would mean that Holmes had started with evidence and *worked down* a chain of known events. That is, if he were given *evidence* and was attempting to *predict* the next murder (of which, would lead to the rendering of additional evidence at a later point in time). However, the novels clearly specify that Holmes would start *from evidence*, and attempt to *put the pieces together*, in an attempt to explain what *had happened*, rather than what *will happen*.

Bayesian Networks can be conducted in the same manner. For example, suppose we

know that it is raining, but we didn't know if there were clouds before the rain came. We can determine the probability that there were clouds before the rain by using Bayes Theorem:

$$
\begin{aligned}
P(C=1|R=1) &= \frac{P(R=1|C=1)P(C=1)}{P(R=1)} \\
&= \frac{P(R=1|C=1)P(C=1)}{P(R=1|C=0)P(C=0)+P(R=1|C=1)P(C=1)} \\
&= \frac{(0.6)(0.4)}{(0.1)(0.6)+(0.6)(0.4)} \\
&= \frac{.24}{0.3} \\
&= 0.80
\end{aligned}
$$

The way in which we would interpret this probability is *not* "there is an 80% probability that it will be cloudy assuming it has rained". Rather, the proper way to think of this is "there is an 80% probability that *it was cloudy* before it rained". That is, abductive reasoning is centered around the premise of speculation of what had happened *prior* to the observation of the evidence. One is not deducing what *will* happen next, but rather, is abducing what had already happed upon the presentation of the evidence. That is, it is speculation of how the evidence came to be rendered.

We can think of deducing as more of a *causal speculation* activity, while abducing is more of a *diagnostic* activity. Such two opposing perspectives makes Bayesian Networks a great candidate for analyzing uncertainties such as supply chain propagation risk. If we had a mapping of possible events that could occur within the supply network that could, but not necessary will, lead to a disruption, then we can leverage Bayesian Networks in two ways: (1) as a diagnostic tool to determine what was the sequence of events that led to the disruption, and, where within the upstream we should expect damage or loss, or (2) as a causal tool to determine the potential consequences of an action upon the observation of evidence. The framework of probabilistic reasoning handles both of these use cases well.

However, the example I have provided is very simplistic in nature. The applicability of Bayesian Networks shine when there are many variables to consider, as well as a conceptualization of those variables. Once the network is built and fully specified, it

can then be leveraged for a variety of purposes (for example, understanding where a problem may have arisen in order to properly identify the "root cause", or understanding the potential consequences of an action, such as running a new type of television advertisement or product launch).

The updating of beliefs are not as simply as I have portrayed above. Applications that leverage Bayesian Networks often have many variables with complex structures. To conduct *belief propagation* in the way we just did would result in an exponential number of equations that would need to be solved, only to simply query the network. When we update beliefs in a procedure described above, the term often used is *inference*, or *belief propagation*. The problem of calculating these probabilities is actually an NP-Hard problem. However, thanks to the advent of technology, super computers, and big data infrastructure, we have the ability to conduct simulation algorithms such as Gibbs sampling and other more general MCMC (Monte-Carlo Markov Chain). Another area of interest in computing these probabilities is that of approximate belief propagation. Such methods use iterative approaches until convergence of the probabilities is attained. As this area of literature is vast,and out of the scope of this dissertation, I will simply direct the reader to the aforecited references.

2.3 Supply Network Structure

As mentioned earlier, supply networks are complex adaptive systems (Carter et al., 2015). Such a structure implies that (1) networks are formed as a result of agents acting dynamically due to consistent changes within the environment and (2) the structure is a manifestation of cumulative behavior of the agents. These two observations alone form the foundation of two distinct, yet related, forms of analysis of these networks: graph-theoretic network theory and social network theory. The graph-theoretic constructs allow the scholar of supply networks to quantify the inherent structural characteristics of the underlying network in a context-free manner. The social network constructs allow the scholar of supply networks to quantify the resulting structural characteristics of various relationships between and among actors, as well as the network itself. Much work has been done in these two areas of study in their application to supply chains. I

will now review this stream of literature that is directly related to the primary inquiry of this dissertation.

2.3.1 The Basic Supply Network Model

As mentioned, (Carter et al., 2015) synthesized the literature of supply networks and arrived at the conclusion that supply networks are inherently complex adaptive systems (Carter et al., 2015). This approach to understanding the supply network has led to a plethora of literature to better understand how networks form or emerge. Before reviewing this literature, it would be worth while to review through the fundamental model of a supply chain itself. Typically, supply networks are modeled as directed graphs (Garvey et al., 2015). While this is a useful tool, the problem in this basic model rests within the criteria of what constitutes a "node" and subsequently an "edge". Much debate has ensued throughout the prior years as to what constitutes a "supply chain".

Peck was among the earlier contributors in the literature to recognize the frustration of multiple definitions. I will ground my definition within her approach, along with my own amendments to it. She offered a thorough breakdown of "the supply chain" by recognizing that it is multi-dimensional and has multiple levels of analysis (Peck, 2005). She offers five levels of analysis of which I will review and amend within the theoretical development section of this dissertation. For now it is important to understand that Peck's contribution of the formation of a framework within which to model supply chains had served as a baseline foundation for supply chain research, offering researcher a tool to carefully describe their indented level of analysis.

In the subsequent sections, I assume that a particular level of analysis has been selected by the researcher. As such, the subsequent discussions on "supply chain structure" are hence based on the assumption that a level of analysis has been selected. However, there is no loss of generality in making this assumption. The literature I present in the next few sections is easily applied to all the aforementioned levels of analysis. Hence, when I say "the node" or "the firm", I am referring to one of the 8 types (of which are fully justified in the theoretical chapter) of nodes described in the framework of levels of analysis for supply networks, with no one in particular to the

discussion at hand. In addition, when I use the words "the link" or "the edge" or "the relationship", again, I am referring to one of the 8 types of connections described in the level of analysis framework, with again, no particular one chosen for discussion of the literature.

2.3.2 The Risk Graph

Some aspects of supply networks that have been previously explored in the literature review of this dissertation are the notions of "risk", "disruption", and "propagation". Such phrases require a conceptualization of the supply network itself to study. For example, simply studying "risk factors" such as what the extant literature does (Qazi, Quigley, & Dickson, 2015; Qazi, Dickson, et al., 2018; Qazi, Quigley, Dickson, & Gaudenzi, 2015; Qazi, Quigley, Dickson, Gaudenzi, & Ekici, 2015; Qazi, Quigley, Dickson, Gaudenzi, & Önsel, 2015; Qazi, Quigley, & Dickson, 2018; Qazi et al., 2017; Lockamy III, 2011; Lockamy III & McCormack, 2010, 2012; Pai et al., 2003) is not sufficient enough to capture the underlying dynamics and interdependencies of risk and it's propagation throughout the network, where such propagations are inherently dependent on the structural design of the network itself. This point was ardently argued in (Garvey et al., 2015). Indeed, they astutely observed that "extant literature in the field of supply chain has been incorporating elements of risk into supply chain models rather than adapting current risk models and placing them within a supply chain setting" (Garvey et al., 2015, pg. 619). The implication of their framework is that supply chain risk should be considered *element-centric*, where "element" refers to a node or link in the supply chain.

As such, (Garvey et al., 2015) framework laid the analytical and conceptual foundation of this idea of the interaction between supply chain risk and supply chain design. Their construction of the "Risk Graph" personifies the appropriate definition of "supply chain risk". Recall that supply chain risk was defined as an interaction between the consequence and likelihood of an event that has the potential to manifest into a supply chain disruption. The definition itself implies that supply chain risk is element-centric.

That is, we cannot simply study "risk factors", but rather, "element-centric risk factors", since the "factors" in one area of the network may impact "factors" in another area and in different manners.

(Garvey et al., 2015) *risk graph* is a probabilistic model that is constructed under a strict set of assumptions (which could of course be relaxed for practical implementation) using the structure of the supply network itself and fusing it together with the variety of supply chain risks that are present within each element of the network (either within a node or within a link). They argue that it is illogical to presume that risk "flows" directly from one risk in a location to another risk in a different location where there is no direct connection between the locations. I admit that this is a shortsighted view of the model, and I will address this shortcoming in the Theoretical Development of this dissertation.

For now, I will briefly review through (Garvey et al., 2015) model, as it serves as the foundation for the analytical methods used in this dissertation. Some of the notation they used in their paper has been altered herein, but I will be clear in meaning here. First, the model of risk propagation measurement begins with a set of nine fundamental assumptions behind the model. They are (Garvey et al., 2015, pg. 621):

1. The entire structure of the supply network is known.

2. All risks considered are modeled as binary random variables.

3. All conceivable risks to a supply network at any location (i.e. at the node or edge level) in the network have been accounted for.

4. Given a set of risks, the causal relationships among the risks are determined in a procedural and objective manner that results in an acyclic directed graph (i.e. given two experts and the same set of risks, they will arrive at the same acyclic, causal structure of the graph).

5. Given a set of risks, the procedure for determining the causal structure of the risks must be based on the structure of the supply network (i.e. given any two locations in the supply network, node or edge, not connected directly in the supply network

in any way, then any risks inherent in one location cannot be causally related to any risks in the other location).

6. The data for all risks and conditional probability tables/distributions can be determined in full (i.e. distributions and parameter estimates for the random variables or the conditional probability tables have already been determined and estimated).

7. The resulting network allows for tractable probabilistic inference.

8. The resulting Bayesian Network constructed using the risks inherent in a supply network using the procedure given below is the "best fit" network to the data.

9. All risks that are dependent on business decisions are static and have only a single distribution (i.e. decisions remain static and result in only a single probability distribution for a risk related to that particular decision).

The model for how risks within a supply network interact is then constructed using a procedure which takes as it's input (1) the structure of the supply network, (2) the collection of all identified risks, (3) the probability distributions between risks, and (4) a procedure that can determine if two risks are causally related (this is somewhat of a corollary of (3)). I will formalize this here.

Let $N = \{x_1, x_2, x_3, ..., x_n\}$ denote the set of nodes in the supply network and let $E = \{e_1, e_2, ..., e_m\}$ denote the set of connections in the supply network. Here, the subscripts on the edges have no particular meaning as to *which* nodes in N are connected. For simplicity, I simply just index them from 1 to m edges. Of course, n represents the number of nodes in the supply network. The supply network is then the union of these two sets $S = N \cup E$.

Now, let $R(z) = \{X_1(z), X_2(z), ..., X_{\phi(z)}(z)\}$ denote the set of risks at location $z \in S$ of the network. Here, it is assumed that $X_i(z)$ represents the i-th risk at location $z \in S$ and is itself a binary random variable with a sample space of $S_{X_i(z)} = \{0, 1\}$. In addition, $\phi(z)$ represents the number of risks at location $z \in S$.

Next, let the set of all random variables (i.e. set of all risks present at all locations)

be denoted as $N_R = \left(\bigcup_{i=1}^{n} R(x_i)\right) \bigcup \left(\bigcup_{j=1}^{m} R(e_j)\right)$. In otherwords,

$$N_R = \{X_1(x_1), ..., X_{\phi(x_1)}(x_1), ..., X_1(x_n), ..., X_{\phi(x_n)}(x_n),$$
$$X_1(e_1), ..., X_{\phi(e_1)}(e_1), ..., X_1(e_m), ..., X_{\phi(e_m)}(e_m)\}$$

So far, we have created a collection of risks in the entire supply network. The next step is to create the collection of connections between these risks. This is done by leveraging the structure of the supply network. (Garvey et al., 2015) specified an algorithm for constructing this collection of connections between the random variables. They make two basic assumptions for the algorithm. First, they define a function for any location in the network which will return the neighbors of that location. I will define it here as:

$$\zeta(z) = \left\{ \begin{array}{ll} N_{G_R}(z), & z \in N \\ \{x_i, x_j\}, & z \in E \wedge z = (x_i, x_j) \end{array} \right\}$$

There is some slight abuse of notation here. Here, G_R will represent the final Risk Graph, this is something we have not yet defined. In addition, while we have indexed the edge set E of the supply network, we have not defined these elements as they usually are defined in graph theoretic language (an ordered pair of vertices, where the first entry represents the tail (source) node and the second entry represents the head (sink) node). Hence, it should be assumed that if
$z = e_k \in E$, then $\exists! i, j \in \{1, n\}$ $s.t.$ $i \neq j, \wedge e_k = (x_i, x_j)$

The next assumption that Garvey makes is that there exists a function, I will denote it $\varphi(X_i(z_1), X_j(z_2))$, that will determine if there is a causal connection between risk $i \in \{1, ..., \phi(z_1)\}$ at location $z_1 \in S$ and risk $j \in \{1, ..., \phi(z_2)\}$ at location $z_2 \in S$. They define this function to be at the discretion of the analyst. I will adjust this function to also determine if the analyst believes that adding such a causal dependency will result in a directed acyclic graph. Recall that Bayesian Networks necessitate the resulting structure to be a DAG. Hence, when I use the notation $X_i(z_1) \Rightarrow X_j(z_2)$, it

will simply represent a query to an "analyst" that will be able to determine if (1) the two risks are somehow related and (2) if the consideration of a causality between the risks will result in a DAG. Hence, for the purpose of their framework, they define it generally. I will simplify their definition here as:

$$\varphi(X_i(z_1), X_j(z_2)) = \left\{ \begin{array}{ll} TRUE, & X_i(z_1) \Rightarrow X_j(z_2) \\ FALSE, & Else \end{array} \right\}$$

Last, Garvey proposes an algorithm to construct the collection of causalities between the risks in the network. I will forgo a description of the algorithm, since it can be expressed more compactly as a mathematical set. To do so, I will define one more function, which will return $TRUE$ if one location in the supply network is a "neighbor" of another location in the supply network when given two risks:

$$\delta(X_i(z_1), X_j(z_2)) = \left\{ \begin{array}{ll} TRUE, & z_1 \in \zeta(z_2) \vee z_2 \in \zeta(z_1) \\ FALSE, & Else \end{array} \right\}$$

Therefore, the edge set in the Risk Graph is defined as:

$$E_R = \{ \ (X_i(z_1), X_j(z_2)) \in N_R \times N_R \mid \varphi(X_i(z_1), X_j(z_2)) \ \wedge \ \delta(X_i(z_1), X_j(z_2))\}$$

The *Risk Graph* $G_R = N_R \cup E_R$ is a directed acyclic graph (DAG) of nodes, which represent binary random variables (which themselves represent risks of disruption), of whose dependencies are determined based on if the risks lie within the same location of the supply network S, or, if the locations of where the risks lie are neighbors of each other in the supply network S, and if the risks are causally related, as determined by some form of "analyst".

2.3.3 Measuring Network Structure

As I mentioned previously, the structure of a supply network is important to understand the various dynamics of supply chains. Past literature has shown that the structure

itself is related to a plethora of various constructs within the supply chain. I will review these later. For the time being, I will review the literature of prior attempts to characterize and measure the supply network structure. Many attempts have been made at this. The logic is usually centered around the ability to operationalize the construct of "supply chain structure" for the ends of relating it to other constructs. Later I will argue that such a pursuit is somewhat misguided since "structure" is too broad of a construct that necessitates the design of sub-constructs, and a different approach should be pursued. There are three broad ways the literature has characterized the structure of supply networks.

The first is through conceptual approaches, which partly rest on the "type" of supply network itself, usually requiring some form of contextual information regarding other constructs such as product category or power. In addition, this category has adopted some constructs of social network theory. There is some overlap between these constructs being more mechanical than conceptual. I will review them under the conceptual, however, since the primary purpose was conceptual by design, and the mechanics are merely operationalizations of those constructs.

The second approach to characterizing supply networks is through a classification approach that is partly dependent on the more mechanical aspects of structure and construct. Classification has been conducted using both conceptual and mechanical definitions of structure, and hence this area of the literature can be thought of as an intersection of the two extremes. We can think of these classifications as the so called "centralization-based" categorization systems. The categories are typically derived from social network theory, and networks within these categories have been seen to exhibit similar characteristics. Last, I will review the more mechanical approaches to characterizing structure. These are the more traditional-based methods to characterize structure. Before the advent of supply-chain, and even social-network theory, these measures were used to solve problems dating as far back as the 18th century. They are still widely used today, and I argue that they are not only sufficient but should be the preferred approach to characterizing supply network structure through the proper construction of the constructs for which these measures reflect.

Context-Dependent Conceptualization of Structure

When I say "Conceptual Characterization", I am referring to a characterization or a classification of the network structure due to theoretical abstract constructs that are not based on inherent and context-free quantitative characteristics. For example, some may classify the "supply network structure" based the types of nodes and how many of those different types of nodes are in the network (i.e. how many raw material sources, how many manufacturer, how many retailers, etc). In any case, there exist many such characterization/classification schemes of which I will now review.

The construct of supply chain structure grew out of those borrowed from the constructs of organizational and enterprise networks (Delfmann & Albers, 2000; M. Dong, 2001; Domínguez Cañizares, 2014). Of course, as the supply chain literature began to show signs of progress, one of the earliest areas of contribution was *mapping the supply chain*, and explaining the various *structural properties* of them. While not explicitly using the term "supply chain" in his research, there is no doubt what Anderson was describing in his study of *business networks*, of which he defined as "set of two or more connected business relationships, in which each exchange relation is between business firms that are conceptualized as collective actors" (Anderson, Håkansson, & Johanson, 1994, pg. 2). His paper had described the nature of a *dyadic relationship* between two firms. His basic model of the dyad comprised of a *focal relationship*, between a supplier and a buyer. The supplier of whom had suppliers themselves and the buyer of whom had buyers themselves.

This early conceptualization of the "supply chain" was used by Anderson to describe the basic structural properties of a *dyadic relationship* that was embedded within a *business network*. He sought to develop network-based constructs that would be useful in elaborating on the nature of dyads. One of the first constructs he proposed was the notion of the *network horizon*, which he defined as "how extended an actor's view of the network is" (Anderson et al., 1994, pg. 4). He argued that this construct is dependent on the actor's perspective within the network. In addition, he defined the notion of the *network context*, which is the part of the network's horizon that an actor concerns

them-self about.

He claimed that this construct is characterized on three primary dimensions: the actors, the activities, and the resources of which the firm is engaged within the "important part" of the horizon. He claims that each actor's context in a network most likely is overlapping with other actor's contexts. Furthermore, given a firm, it's network context and the network itself, Anderson argues that a firm also establishes its *network identity*, which he defined as "the perceived attractiveness (or repulsiveness) of a firm as an exchange partner due to its unique set of connected relations with other firms, links to their activities, and ties with their resources" (Anderson et al., 1994, pg. 4). Furthermore, Anderson argues that network identities too are dependent on the three aforementioned dimensions.

Last, he attempts to characterize how *embedded* a dyadic relationship is within the business network by defining what he referred to as *relationship connectedness*. These constructs are distinct from the graph theoretic term *connectedness*. The idea behind his contribution of these constructs was to eventually determine how "well connected" a dyadic relationship is with respect to the network. This is inherently a contextual issue of the relationship. The two constructs he offered to help describe this notion of "embeddedness" were *anticipated constructive effects* and *anticipated deleterious effects* on network identity. These constructs essentially characterize the firm's perception of other firms in which it is part of a dyad and how these dyads affect it's identity in either a positive or negative way.

These early characterizations of network structure laid a foundation for the early scholars of the supply chain literature to design frameworks from where a supply network structure can be characterized. Of course one of the most prominent contributions was from Lambert, Cooper, and Pagh, who offered one of the first characterizations of the structural dimensions of a supply network (Lambert et al., 1998). They proposed that the supply network be characterized by the dimensions of the *horizontal structure*, the *vertical structure* and last the *horizontal position* of the focal firm. The horizontal structure is defined as the number of tiers across the supply chain while the vertical structure is defined as the number of suppliers/customers within each tier. Last, the

horizontal position refers to how near or far the focal firm is from the initial source of supply or the ultimate consumer.

Around the same time Lambert's model of "supply chain structure" was being developed, Fisher had suggested that supply chains should be differentiated by innovative versus functional products (Fisher, 1997). Lamming built on this work by empirically validating Fisher's model and also suggesting that supply chains should be categorized based on two primary dimensions: the innovativeness of the product and the complexity of the product. They argue in grounded research and from empirical observations that supply chains must fall within one of the four categories determined by these two dimensions (Lamming, Johnsen, Zheng, & Harland, 2000). They followed this work with a subsequent study where they further elaborated on their dimensions of firms being "dynamic" and the degree of "influence" of which they possess (C. M. Harland, Lamming, Zheng, & Johnsen, 2001).

Another proposal for describing the structure of a supply network was characterizing different attributes that firms possess as well as the characterization of their inter-connections being either financial, informational or material (Swaminathan et al., 1998). A similar idea was proposed by Choi, who suggested that supply networks should be studied through the lens of a complex adaptive system. In addition, he argues that the characterization of a network overall could be based on three primary dimensions: internal mechanisms, environment and co-evolution. While his characterizations were still fairy conceptual, he did mention the that the constructs of complexity and connectedness are most likely involved in the determination of how a supply network evolves (Choi, Dooley, & Rungtusanatham, 2001) as well as having the potential to explain firm attributes and behaviors. Similar arguments were also suggested in Harland's work (C. Harland, Zheng, Johnsen, & Lamming, 2004).

Despite these early conceptualizations of characterizing the structure of a supply network, most of these were inherently reliant on qualitative attributes of the network such as product type of firm attributes. Meyr was among the first in the literature to suggest that networks can be characterized along two primary dimensions: functional characteristics and structural characteristics. The "functional" comprised of similar

attributes that prior work had pointed out. However, the "structural" dimension was among the first in the literature to describe the "structure" of the network from more of a quantitative perspective, by suggesting this dimension could be explained by variables such as power imbalances and degree of globalization(Meyr & Stadtler, 2005).

Similarly, Vereecke argued that firms can be characterized within their network by not only observing their functional properties but also their structural properties in so far as where they lie within their network (Vereecke, Van Dierdonck, & De Meyer, 2006). Likewise, Singh had argued that supply chains should be characterized by their network topology, flows of information and material, governance between key network partners, and a "value structure" of the product (Singh Srai & Gregory, 2008). Sentiments were the same in other research that suggested characterization by link type as well as different forms of "complexity" (Bode & Wagner, 2015; Pedroso & Nakano, 2009). Last, others have characterized network structure based on firm strategy (Kilubi, 2016), embeddeness within the network (Yan, Choi, Kim, & Yang, 2015), as well as using industry-specific attributes for such characterizations of structure (MacCarthy & Jayarathne, 2013).

Partial-Context-Free and Context-Dependent Classification and Taxonomy

Conceptual attributes of network structure help simplify the complex nature of networks. As with any abstract construct, they help us understand the intuitive nature of something that is abstract, while helping us to filter away "garbage" to better understand the underlying attribute (Churchill Jr, 1979). There is little doubt that conceptual characterizations of network structure aid us in understanding it's abstract nature. So too, however, does classification.

While conceptualization of complex attributes of an entity help us gain an intuitive sense of the underlying attribute, it does little in the way of understanding how "similar" or "different" the entity is from another. In our case, that of network structure, we must ask, "how do we distinguish between one entity from another?". That is, not only would we seek to be able to identify a supply network, based on it's structure, as "this" or of "that form", but as a second criteria, be able to distinguish it from other

individuals (networks) (Sneath, 1957b, 1957a). Conceptualization aids in our thinking for the former, but not so much in the latter (depending on the definitions and variability of said conceptualizations). Hence, an alternative approach to characterizing network structure would be through a proper classification (i.e. a typology or taxonomy).

However, how might one construct such a classification system? As we have already seen from a conceptual point of view, there are primarily two dimensions in which the extant literature seems to strive towards classification of supply networks: structure and function (the former sounds a bit recursive, but is salient none-the-less). The taxonomy or classification is not only based on our observations of the entities we seek to organize, but also is based on our planned goals for the taxonomy. Indeed, a "classification is greatly influenced by the purpose for which it is devised" (Sneath, 1957b, pg. 184). A logistics manager would hardly find it useful to classify networks based on corporate strategy, and a supply chain manager may find it hardly useful to classify based on mathematics. This would explain the vast number of attempts in the extant literature to design "The Supply Chain Taxonomy." Despite this, I will review through a sample of them now to cover what has been proposed in the past as motivation for the final section in this chapter.

Some of the notions of "types of supply networks" grew out of Lambert's work. When he defined Supply Chain Structure as having a horizontal dimension, a vertical dimension, and a horizontal position, some academics felt it natural to extend this towards a classification of supply networks (vertical networks vs horizontal networks) (Lambert et al., 1998). Choi extended this idea, despite reducing these three dimensions to a single network characteristic in his own framework (of which he termed as "complexity", the other two respectively being "formalization" and "centralization"). Indeed, "in the supply network, the measurement of complexity in terms of these three variables [vertical, horizontal, distance] becomes rather straightforward, once a complete map of the supply network has been obtained. It would simply entail counting the number of tiers (vertical complexity), the number of suppliers in each tier (horizontal complexity), and the average distance between two firms engaged in buying and supplying (spatial complexity)"(Choi & Hong, 2002, pg. 471).

What is interesting is that around the time that Lambert and Choi, respectively, published their work on supply chain structure from a more theoretical lens, another stream of literature in the supply chain structure branch was at it's infancy, despite originating in a entirely distinct area of study. In 2001, the now-famous paper by Strogatz, was published, describing the general idea of "small-world networks", and characterizing this class of complex networks as one where some notion of order can be found in disorder (Strogatz, 2001). It did not take long for academics in the supply chain literature to adopt Strogatz ideas, and soon an entire new form of classification was born.

Rather than classify networks based on the relation of a firm to it's network, or it's level of "embeddeddness", the structural classification approach, based on graph-theoretic measures, was beginning to show signs of progress in the literature in their applications to supply chain management. Indeed, Pathak offered a taxonomy of supply chain structures as having no structure, centralized, linear, flat, hierarchical, federated or of starburst design. He offered clear graph-theoretic criteria for each of these groups of networks, as well as example industries in where one might likely find them (Pathak, Dilts, & Mahadevan, 2009). In addition, based on the structure, he offered propositions in regards to a firm's survival and ability to compete in different types of networks.

The more "graph-theoretic" approach to classifying networks was also extended, when it was noticed that social network theory (an extension of graph theory) could be applied to supply networks to help explain a wide array of phenomena. Such theories leveraged the mathematical properties of the graphs, but only doing so in a way that allowed for meaningful classifications (where the criteria for classification was based on the mathematics of the graph) (Capaldo & Giannoccaro, 2015b; Y. Kim, Choi, Yan, & Dooley, 2011; Galaskiewicz, 2011). In addition, the approach in this stream has also been applied to risk analysis (Ledwoch, Yasarcan, & Brintrup, 2018; Basole & Bellamy, 2014b) as well as trust and firm performance (Capaldo & Giannoccaro, 2015a).

This stream of literature had eventually matured to a point where a standard classification of types of networks were discovered (Pathak et al., 2009; Basole & Bellamy,

2014a; Y. Kim et al., 2015). Pathak had categorized different networks based on graph-theoretic metrics such as maximum depth, degree, if they network did not have any directed cycles, and if the the network resembled rings or trees (Pathak et al., 2009). He argued that there were four major categories of networks: No Structure, Centralized, Linear, and Hierarchical. On the other hand, Capaldo, who was seeking to determine if structure affected trust, had classified networks similarly to how social scientists classify them: Erdos-Renyi random, Ring, Small World, Block Diagonal, Preferential attachment, scale-free-like, hierarchical, diagonal, centralized, and dependent (Capaldo & Giannoccaro, 2015a). Rivkin also used a similar classification of networks in his research to better understand the structure of complex systems, where his classification comprised of random, local,small world, block-diagonal, preferential attachment, scale-free, centralized, hierarchical, diagonal, dependent. (Rivkin & Siggelkow, 2007)

The second stream of literature for supply network classification is one that originates from Lambert and Choi (Choi & Hong, 2002; Lambert et al., 1998). Indeed, they recognized that the structure of a supply network has two important perspectives: the micro and macro, one of which grew out of the other in the literature. From the macro-perspective, the notions of vertical and horizontal size were defined. Choi generalized this notion a bit further. He understood that Lambert's three dimensions intuitively represent how *large* the network is and where within the network a focal firm lies. Given this notion of magnitude, together with it's size increasing with actors dynamically changing the network structure, he defined these three concepts, when taken together, as "complexity". Furthermore, Choi understood at the time that looking only at dyadic relationships was myopic, and that if they sought to advance in their performance, they would need to look outward toward the horizon, on the macro level (Choi & Kim, 2008). He referred to this concept of being "in the network" as *structural embededness*. That is, where does the firm lie with respect to the rest of the network?

Choi's notion of structural embeddedness is a generalization of Lambert's notion of "horizontal position". It sought to emphasize not only how well connected firms were relative to their overall dyadic relations, but also, how "deep" within a network a firm

lies. Other works in the literature stream soon adopted this idea, with some hypothesizing that embeddedness was related to a firm's performance (Choi & Kim, 2008; Bernardes, 2010). In addition, some attempted to fuse the works of Choi and Lambert's notion of structural characterization of a supply network with that of the more traditional functional characteristics mentioned earlier (Giannakis & Croom, 2004).

In addition, some sought to understand embeddeness within a larger scope of research such as supply chain mapping and it's relation to supply chain performance (Iyengar, Rao, & Goldsby, 2012; Yu, Suojapelto, Hallikas, & Tang, 2008). Innovation was also explored within the context of network embedesness (Ahuja, Soda, & Zaheer, 2012; Schilling & Phelps, 2007; M. C. Dong, Liu, Yu, & Zheng, 2015). Ironically, however, Choi's initial work on more conceptual structural metrics of a supply network began to turn more to the quantitative side. Indeed, Choi himself had contributed to this area by trying to further quantify the construct of "embeddness" (Y. Kim et al., 2011) via graph-theoretic measures such as centrality and distance-based metrics. In addition, some had continued to expand Choi's contributions by determining how structural embededness can be measured and simultaneously be understood in connection with other the processes used to map the supply chain (Kito, Brintrup, New, & Reed-Tsochas, 2014; Carnovale, 2014). This area of "embeddness-based classification" grew ever so closer to the more quantitative social network end, and it could be argued that the two forms of classification have converged in ideas over the past years.

The second perspective that Choi took was that from the micro perspective. He realized that one way in which we can understand a firm's position within the larger network is if we also understand where it lies within it's local network. In so doing, Choi applied the concept of *triads* to supply network structure theory. He argued that triads form "the fundamental building block of networks" (Choi & Wu, 2009, pg. 10). In addition, it brought to light yet another concept from social network theory of which soon was applied to supply network study: structural holes. A structural hole is "a special type of triad occurs frequently in the supply network, wherein a buyer keeps suppliers apart in order to engage them in competition" (Choi & Wu, 2009, pg. 11). Hence, Choi helped bring focus to the study of triads in the SCM literature of the day.

The construct of the triad allowed research to flourish in the area of structural studies of supply networks. Dubois had posited that triads can explain certain sourcing policies as well as competitive behavior (Dubois & Fredriksson, 2008), and Wu found that certain forms of triads impact supplier performance (Z. Wu, Choi, & Rungtusanatham, 2010). Not only are triads are the building blocks of supply networks, but it could be argued that they are the building blocks to yet another important construct taken from social network theory: the ego-network.

An *ego-network* is the network of individuals that are neighbors (they are a buyer or a supplier) of a firm in the supply network, along with their respective links between them (if they exist). The research in dyads and triads eventually grew more outward to the study of ego-networks. Borgotti had suggested to borrow concepts and measures from social network theory to help quantify ego-networks and their structural characteristics. In addition, Carnovale had found that ego networks can explain the outcome of potential joint-venture formations (Carnovale & Yeniyurt, 2014).

Hence, to summarize, the literature calls for a vast and varying degree of different taxonomies for supply networks. The extant literature seems to suggest classification in one of two manners: through the dyad/triad/ego-network perspective and through the more graph-theoretic perspective. However, my review on this matter is limited, and intentionally so. For a more in-depth review of the varying different types of network classification, see (Bellamy & Basole, 2013).

Graph-Theoretic Based Characterization

Thus far, I have ventured into two primary ways on how supply chain structure can be characterized: (1) through conceptual methods and context-dependent information, or (2) through the classification approach. As mentioned, for the classification approach, researchers have found that certain types of networks tend to share similar types of graph-theoretic properties, namely in their adjacency matrices (Rivkin & Siggelkow, 2007; Pathak et al., 2009). However, simply just categorizing networks is not the magic hammer that can fix all problems. Indeed, as mentioned earlier, classification systems are often designed for a well-designed goal, objective and problem (Sneath, 1957b,

1957a). Furthermore, if classification of graphs themselves are based on the metrics (Watts & Strogatz, 1998), then it begs to ask, why are we bothering to classify in the first place if we have the raw data? (Answer: it depends on the problem).

Unlike classification systems or conceptual-based characterization of networks, graph-theory provides the perfect tool box to objectively be able to understand a network's "structure". In the remainder of this section, I will offer a review of the literature that has taken this third approach to characterizing network structure. In addition, I will summarize some of the most important aspects of graph theory, and it's application to understanding structural characteristics.

Graph theory has it's roots in practical problem solving. The motivation for the theory began with Leonhard Euler and his now popular Seven Bridges of Knigsberg Problem (Euler, 1736). The problem that he was seeking to solve was to determine if there was a walk around the seven bridges of Knigsberg, of which connected two parts of mainland to two islands, in such a way that he will not have to cross the same bridge twice. He later found that there had existed no such solution, and hence it was deemed impossible.

Euler's approach to the problem, however, had motivated the study of a new field of mathematics: namely Graph Theory (and, on a mildly related topic, it also motivated the field of Topology). Since then, Graph Theory has been applied to a wide range of different problems whose solutions were not as evident had it not been for the model and study of a graph. Throughout the years since Euler, many have described the properties of graphs by constructing measures which takes into account *the whole network*, as well as properties concerning the nodes and edges.

Arc/Node-Level Measures

As I defined in an earlier section, a *graph* is a set that is composed of two other sets, one being the collection of *vertices* (also known as nodes), and the other being the collection of *edges* (also known as arcs, links). Once a graph is constructed, one often seeks to describe the attributes of the graph. There is a plethora and a vast diversity of different measures that one can calculate to describe these attributes. On a node/arc level, the following are common measures used to gain a clearer understanding of that

node/arc's attribute within the graph and their influence.

1. Betweenness Centrality
2. Closeness Centrality
3. Eigenvector Centrality
4. Degree Centrality
5. Eccentricity
6. Min Distance from Center
7. Max Distance from Center
8. Average Shortest Path
9. Density
10. Number of Triads
11. Number of Structural Holes
12. k-Ego Connectivity
13. Coreness
14. Clustering Coefficient

Network-Level Measures

There are quite a few measures that capture some of the structural characteristics at the higher network-level of analysis of which I have summarized below. Many of these characteristics fully explain the graph's "structure". In addition, some of them are averages or centralized versions of their node counter-parts. When a scholar of graphs studies the network-level structure, what they often encounter is a *distribution* of the node-level values, which is contrasted with the node-level metrics (no such issue arises in that instance). As such, some have proposed methods for combining these values so as to gain an intuition for the "network's structure". The primary ways in which

to accomplish this within graph-theory are through the methods of centralization of a measure and the average of a measure.

When one computes measures that are at the network level of analysis, there are two general classes of such metrics. The first collection comprises of those where a simple addition or count is necessary to compute the measure. For example, the "node connectivity" is a network-level metric, and is computed by finding the minimum number of nodes that must be removed from the graph in order to disconnect it. Such a calculation was not inherently an amalgamation, combination, or more generally a function of node or edge level attributes. That is, these types of measures can readily be computed without first computing node/edge-level attributes.

On the other hand, the nodes and edges within a network themselves have properties. One can describe the structural characteristics of the overall network by also summarizing the node/edge level characteristics. However, even this can be done in a multitude of ways. One of course take a more traditional statistical approach by calculating the various moments of the data (the mean, the variance, the skewness, etc). A tangential approach that is more aligned with graph theory is to use *Freeman's Centralization*.

This metric is inherently different from the traditional statistical approach in two primary ways. First, the Freeman Centralization seeks to characterize how *central* a network is by finding the average distance of each element in the network (typically only the nodes) from the maximum of such a distance. That is, unlike the mean, which seeks to find a point of balance within the data with respect to different, and unlike variances which seeks to find the average squared distance from the center (as measured by the mean), the centralization measures the average distance in the network from the highest such values of centrality.

There are commonly two forms of centralization: *normalized* and *non-normalized*. A normalized centralization finds the average distance of a node in the graph from the maximum value of a centrality measure and divides this by the maximum distance of a node in the graph from the maximum value of the same centrality measure among all other graphs of the same size. That is, the measure is dependent on looking towards

other graphs of the same size. The *non-normalized* approach does not conduct this division, and typically keeps the value as is.

More specifically, for a node $n \in N$, if we have a centrality measure , (a function $C : N \longrightarrow \mathbb{R}$), then the *Freeman Centralization* for the network as a whole would be:

$$Centralization(C(n)) = \frac{\sum_{n \in N} \left(\max_{k \in N} \{C(k)\} - C(n) \right)}{\max_{N_j \in N^*} \left\{ \sum_{n \in N_j} \left(\max_{k \in n} \{C(k)\} - C(n) \right) \right\}}$$

where N^* represents the collection of all graphs of size $|N|$. This is the normalized version of the metric. The non-normalized version of the metric would be:

$$Centralization(C(n)) = \frac{1}{|N|} \sum_{n \in N} \left(\max_{k \in N} \{C(k)\} - C(n) \right)$$

Again, I enumerate a list of known network-level measures that mathematically characterize the network structure. I will review through these and the node/arc level in a later chapter.

1. Number of Nodes/Arcs
2. Edge Connectivity
3. Node Connectivity
4. Diameter
5. Radius
6. Girth
7. Circumference
8. Clique Number
9. Number of Largest Cliques
10. Average Clique Size

11. Min Distance from Center
12. Max Distance from Center
13. Density
14. Number of Triads
15. Number of Structural Holes
16. Average/Centralization Betweenness Centrality
17. Average/Centralization Closeness Centrality
18. Average/Centralization Eigenvector Centrality
19. Average/Centralization Degree Centrality
20. Average/Centralization Eccentricity
21. Average/Centralization Shortest Path
22. Average/Centralization Constraint
23. Average/Centralization k-Ego Connectivity
24. Average/Centralization Coreness
25. Average/Centralization Clustering Coefficient

2.3.4 Network Structure Dynamics in Supply Chains

So far I have discussed a bit of the background in the literature for prior attempts to characterize the "supply chain structure". Some methods, as I mentioned earlier, are more contextual in nature. For example, we may classify a supply chain into a"Banking Supply Chain" if it's primary business is banking. We may classify a toy manufacture as a "manufacturing supply chain", so on and so forth. These may be insightful for managerial purposes, however, the lack of concern of so called objective structure inherent in the graph itself is myopic to not only strategic management, but also towards other constructs pertaining to supply chain.

Those in the extant literature of the recent prior few years seem to be converging onto more and more of a consensus on a proper characterization of supply networks. They often propose that a fusion between the contextual-factors and the graph-theoretic constructs. In addition, this lends a very helpful hand for firms attempting to understand, and possibly extend, beyond their horizons. Some in the extant literature have had it as their research goal to "map" the entire supply network of a focal firm. Such an endeavor has led to a steady-increase in Agent-Based Modeling work,evolutionary algorithm design, and more generally, simulation approaches applied to and with the theory of complex adaptive systems (Nair & Vidal, 2011; Pathak, Day, Nair, Sawaya, & Kristal, 2007). The goal of these researchers, using these methods, was to simulate a real sequence of events within a supply network based on previous and empirically sound attributes. The researcher would then keep track of the evolution of the supply network and gain a further understanding of the transcendence of the network itself into a particular pattern or state (Snyder et al., 2006).

Indeed, much work in this area of the literature and has already been conducted for a variety of reasons such as studying ordering strategies (Lin & Wang, 2011), how firms handle risk and resiliency (K. Zhao, Kumar, Harrison, & Yen, 2011), their ability to design robust networks (Zokaee et al., 2017), and to ascertain just simply the structural properties of the supply network that changes over time as it's complexity increases (Orenstein, 2016). The network design of a supply chain has been shown in the past to affect a variety of different constructs in supply chain. Our anecdotes aside, supply chain structure has been show to affect performance (Choi et al., 2001), innovation (Choi et al., 2001), trust (Ireland & Webb, 2007), information and knowledge sharing (Kotabe, Martin, & Domoto, 2003), and risk and disruption (Han & Shin, 2016)

2.3.5 Why Bother to Classify?

A simple question that comes to mind when studying most of the aforementioned topics is, "why bother to classify supply networks?" Graph-theory allows us access to a plethora of numerical equations that help us further quantify the structure. Again, while classification and context-dependent characterizations may help managers with certain

tasks, they lack the ability to fully understand the dynamics between the inherent structure of the network and whatever we are attempting to determine the structure affects. That is, by using classification methods to study the effects on a variety of dependent variables, we are, in essence, "throwing away" good data.

I hence contend that a more appropriate manner to study the affects of structure on whatever our intended topic at hand is, is to first characterize our networks using the variety of measures discussed above. Naturally, we can expect a lot of correlation between groups of these measures. However, simply categorizing the networks and pigeonholing them into groups, as what has been conducted in past research with structural-based research, would essentially be intentionally increasing the variance of the dependent variable for no apparent reason. Further, it would be more difficult to isolate out effects on dependent variables by categorizing. Hence, one of the underlying assumptions behind this dissertation is that categorizing for the purpose of analyzing cause and effect relationships on supply network metrics, when no such categorization is necessary, would essentially be throwing away good information, and hence, it will be avoided.

2.4 Current Research Gaps in the Literature

Given the literature just reviewed, I have identified a few gaps in the current stream of literature as it pertains to structure and risk, as well as some tangential topics as well. In this section, I will discuss some of these gaps, and motivate the final research questions at hand.

2.4.1 Context-Free Network Structure Conceptualization

As the literature review illustrates, there are many perspectives on what constitutes a "network structure". This author's perspective is on the canonical and context-free perspective. That is, given a network, any network (a computer network, a distribution network, a supply network, a social network, etc), can we describe the inherent

structure of the network, without relying on a particular context (i.e. a definition, characterization, or classification of the specific nodes and links), and/or, without engulfing ourselves in a sea of mathematical measures, that are either seemingly unrelated or overly duplicated? I argue in the affirmative, and not only so, but also argue that such a characterization is of extreme importance to understanding a variety of supply chain construct dynamics, many of which have already been studied from other structural perspectives.

The literature review illustrates that there have been multiple attempts to characterize the supply network structure. I have organized this literature into three primary areas of discussion that seem to be salient: Context-Dependent Conceptualization of Structure, Partial-Context-Free and Context-Dependent Classification and Taxonomy, and Graph-Theoretic Based Characterization. The literature that has specifically focused on the topic of characterization has primarily been within the context-dependent conceptualization of structure. While these may be useful for isolating supply chain specific construct dynamics, there is little discussion on how the canonical structure, as Choi defines it, is related to the various aforementioned supply chain constructs that have been studied. For example, a researcher may seek to understand how "supply chain structure" impacts the continuance of a buyer-supplier relationship. Their characterization of "the network structure" may be via a classification of nodes and links (such as having types of nodes like a manufacturer, a retailer, etc, and types of links such as information, material, financial). Again, this may be of interest with respect to that specific characterization of the network, but speaks little to the relation between the canonical structure of the network and the constructs under study.

A recent advancement has been the convergence of literature towards the classification of supply network structures, and the design of structure-based taxonomies, which appear to be converging towards a consensus within the literature. However, as I have previously argued, classification based systems are essentially throwing away good data. In addition, while theoretically interesting, it is quite difficult to justify their usefulness within practice. Rarely would a manager know the structure of the network they

lie within. Hence, understanding weather or not a scale-free based network is more/-less "risky" than a small-world network may be an interesting thought experiment, but that is about the only usefulness to practitioners that exist within these results of recent research.

Last is the topic of graph theoretic-based characterizations. I have argued that these indeed have been useful in a plethora of studies in being able to isolate the effects that the "canonical structure" has on certain supply chain constructs such as power, trust, joint-venture formation, collaboration, etc. However, the primary drawback to this area of research is a lack of a conceptual characterization of "structure", and a lack of generalizability of the results. For example, if a researcher happens to find that "closeness centrality has a positive relationship with power", for example, is a relationship not between a construct and another construct, but rather an *observable* and another construct. Again, while interesting, the results found in this area of research lack generality, and are heavily reliant on the graph-theoretic measures, rather than on the constructs of which the measures reflect.

To the best of my knowledge, given the review of the supply network literature, the complex adaptive systems literature, the graph-theoretic literature and the social-network literature, there appears to be a lack of a conceptual framework that would describe context-free network structural characteristics. For example, relating back to the aforementioned example, it would be "more interesting" to understand the general relationship between a general construct of "centrality" and "power", rather than a specific relationship between the observable "closeness centrality" (a more specific form of "centrality") and "power". Interestingly enough, such a framework seems to not exist in the literature.

The reason for the need of such a framework is to be able to transcend an intuitive conceptualization of "structure" without having to rely on (1) context or (2) mathematical measures. There exist many graph-theoretic measures to understand "structure". Many of them, however, are correlated and related, yet distinct. I argue that many of them are actually observables of a collection of more general structural constructs, part of a larger framework of a "Theory of the Network". The implications of identifying

such a framework of constructs would be immense, not only for supply chain theory, but also, for any science that leverages network design. I hence contend that this is a major gap in the extant literature.

2.4.2 Network-Based Hierarchical Levels of Structural Analysis

Yet another gap within the literature is that of a conceptual network-structure "level-of-analysis". That is, when studying networks, what is the unit of analysis we seek to further explain? As far as I can extract from the extant literature, there seems to only be a small taxonomy of varying levels of analysis within a network, namely: global—level (the entire network), ego—level (with respect to a node), triad—level, dyad—level, and node level. Some instances do involve the notion of a "core", but these are much more mathematical in nature, and while may be an interesting topic to further explore at a later point in time, still lacks the justification to be placed in a taxonomy of level of analysis of networks.

I submit here, however, that a network inherently has more levels of analysis than what currently exists in the aforementioned taxonomy. More specifically, I argue that the construct of the ego network can be further generalized outward towards what I define as the k-ego-network. I will define this construct more thoroughly in the theoretical portion of the dissertation. For now, we can think of this construct as the ego-network *of* the ego network *of* ... the ego network *of* a node (k-times). For example, a 2-ego-network of a node would be the induced subgraph of the union of nodes in the respective ego-networks of the nodes in the primary ego-network of interest. This notion will be illustrated further below in the theoretical development. However, I argue the need for such a construct in order to understand how relationships between a variety of supply network constructs may differ across the differing levels of analysis, which may predict a class of new managerial-based problems for researchers to answer.

2.4.3 Risk Propagation

In this dissertation, I will loosely use the phrase "supply chain risk" or "supply network risk". It should be noted, however, given the literature review above, that there is a

great distinction between a supply chain risk and a supply chain risk propagation. These two constructs are indeed related, yet, very distinct. From what I can gather in the extant literature, there seems to lack a proper conceptual line of demarcation between these constructs. Indeed, many researchers seem to conflate them, and even define them as equivalent phenomena.

It is important, however, to understand that these are indeed distinct constructs. I feel that a current gap in the literature is a proper line of demarcation in the definition of these constructs. I will motivate such definitions in the theoretical development. It should be very clearly stated that in this dissertation, the primary and even tangential topic of concern is NOT on supply network risk, but rather, on supply network risk *propagation*. However, in the interest of brevity, the reader should be aware that when I refer to *supply chain risk*, I will be doing so synonymously with *supply chain risk propagation*, since the context of this dissertation is clear. The reader should, however, be aware that these are very distinct conceptualizations, as I will outline within the theoretical portion of this dissertation.

2.4.4 Intersection of Risk Propagation and Network Structure

As far as I have researched, there is very little research at the intersection of risk and network structure. As I reviewed earlier, much of the risk literature has focused on measuring risk from a subjective standpoint, or focusing on only a single location (Heckmann et al., 2015; Olson & Wu, 2011). The measurements used in much of this area do not capture the interdependencies between risks. The studies that do attempt to measure risk propagation (Qazi, Quigley, & Dickson, 2015; Qazi, Dickson, et al., 2018; Qazi, Quigley, Dickson, & Gaudenzi, 2015; Qazi, Quigley, Dickson, Gaudenzi, & Ekici, 2015; Qazi, Quigley, Dickson, Gaudenzi, & Önsel, 2015; Qazi, Quigley, & Dickson, 2018; Qazi et al., 2017; Lockamy III, 2011; Lockamy III & McCormack, 2010, 2012; Pai et al., 2003) have not ventured yet into studying how the structure of a network affects the propagation, with the exception of a notable few (Y. Kim et al., 2015). The studies that do attempt to ascertain this connection, however, have only focused on them using a classification system, rather than using the underlying metrics

that capture the structure of the supply network. Hence, a large gap in the current literature is the intersection of risk propagation and the underlying network structure.

2.4.5 Supply Chain Horizon and Risk Propagation

In addition to propagation and risk, the extant literature that lies at the intersection of supply chain horizon and risk propagation is mostly strategic (Busse, 2016; Barratt & Oke, 2007; Wei & Wang, 2010). Few attempts have been made to quantify the notion of "the horizon". In addition to this lack of quantification is the lack of studying it's impacts on propagation, containment of propagation, and among other related topics. Indeed, (Carter et al., 2015) contends that practically no research is in existence on topics pertaining to the supply chain that is outside the visibility of the horizon, let alone those aspects and their connection with risk propagation (Carter et al., 2015). Therefore, a useful first step towards understanding the connection between "the horizon" and risk propagation would be to first quantify "horizon".

2.4.6 Hierarchical Levels of Structural Analysis and Risk

2-Ego Network Risk

As mentioned, there is little research that attempts to quantify the "visibility" of the horizon. One approach to do so is by looking at a firm's ego-network. However, this is not enough. Ego-networks themselves are embedded in a larger network, as I have discussed in the above section. Hence, if we had a notion of "continually embedded ego-networks", this may serve useful in our quest in not only understanding propagation, but propagation that is outside the visibility of the horizon. I there informally (later to be formalized) define the following:

A *k-ego network* of a node or edge is the induced subgraph of the nodes in the union of each node's own *(k-1)-ego network*. Therefore, this is defined recursively. A 0-ego network would be the node or edge in a supply network. A 1-ego network would be the ego network as prior literature has discussed. A 2-ego network would be the induced subgraph on the union of each node's own ego network in the first ego network, and so

on.

The notion of the *k-ego network* allows the researcher a way to quantify "the horizon". The logic being that a firm would be close with it's first ego network, but as we go outwards, the focal firm looses information of the other actors in the subsequent higher-order ego networks. However, it can be speculated that the first and second order ego-networks is sufficient information for the firm. It would be difficult to believe that a firm is aware of their supplier's supplier's supplier, and their supplier's customers' customers (or supplier's).

While obviously an entire enumeration is not possible, it would be a fair assumption to presume that the focal firm, dependent on some characteristics of itself of course, may have sufficient knowledge up to and including the 2-Ego Network level. As this is a new concept of "visibility", there obviously has not been any research conducted on it's relation to supply network risk propagation. Hence, this is yet another possible area of exploration

Ego-Network Risk

While I have just introduced the concept of the k-ego network, the 1-ego-network has been known to scholars for quite some time. Shocking, however, is the lack of research in it's connection to risk. As far as this author is concerned, there exists no literature exploring the link between ego-network and risk, let alone risk propagation. Assuming this is true, then there exists great opportunity to study how the structure of ego-networks impact the overall ego-network's risk. In addition, given the embedded structure of the definition of k-ego networks, it would be interesting to determine if risk "trickles down" from one level of analysis to another.

Node and Arc-Level Risk

As I have demonstrated earlier, much work has been conducted in the past to study how robust, resilient and "at risk" a firm is within a supply network. What is currently missing from the extant literature, however, is the study of structure and arc-level risk. Few studies have been conducted on this potential connection, and most concern

themselves over transportation risk (Linh, Amer, Lee, & Phuc, 2018; Baharmand et al., 2017; C. Li, Ren, & Wang, 2016). What would be an interesting development in the literature would be to determine if node-level risks are affected differently than arc-level risks. In addition, the risk of propagation and the structural characteristics of node/arc would indeed be a useful contribution if such a connection is found.

2.4.7 Research Questions

Given the literature review and the identified gaps in the extant literature, I therefore submit the following research questions that this dissertation seeks to address:

1. What are the underlying conceptual constructs of a network structure that current graph-theoretic measures would reflect?
2. How does 2-ego network structure differ from 1-ego network structure?
3. What is the impact of 2-ego network structure on 2-ego network risk?
4. What is the impact of 1-ego network structure on 1-ego network risk?
5. What is the impact of Node Structure on Node risk?
6. What is the impact of Arc Structure on Arc Risk?
7. Does the relationship between network structure and network risk depend on the $k - ego$ level?
8. How does 2-ego network risk impact 1-ego network risk?
9. How does 1-ego network risk impact Node/Arc network risk?

Chapter 3

Theory Development

In this chapter, I will lay the theoretical foundation upon which this research is built. First, I will offer a discussion on the general study of supply networks. I will offer three points of contention, namely, that (1)specific to supply networks, a proper taxonomy of supply network level-of-analysis needs to be clearly defined, (2) more generally to any network structure, there exist a greater depth to the current taxonomy of general network-based levels of analysis, and (3) the canonical structure of a network needs to be characterized through the use of a general and conceptual framework of constructs.

I will first argue that the current taxonomies of levels of analysis within the supply chain literature is lacking, and even to certain degree, ambiguous. I will identify a workable taxonomy that is currently salient in the extant literature, but argue that it is missing what I consider very important amendments for the sake of theoretical completeness, as well as offer a theoretical foundation upon which I will rest other arguments for tangential topics within this dissertation.

After, I will argue that the level of analysis from a pure canonical network perspective is also incomplete. Thus far, most researchers study relationships within networks from either a "global perspective" or a "local perspective". I argue that studying relationships within these two extremes is either impractical (in the case of the former) or myopic (in the case of the latter). As such, I argue that a general network structure actually has a much more refined taxonomy of levels-of-analysis. I will argue that the introduction of a construct of which I have defined as the k-ego network is a solution to the problem of understanding relationships in-between these two extremes.

From the perspective of characterizing a canonical network structure from a conceptual point of view, I will then establish a theoretical framework upon which I argue

network structures rest upon. Throughout my literature review, I have found, shockingly, that no attempts have been made to synthesize the graph-theoretic and social-network-theoretic measures into a handful of conceptual dimensions (constructs) of the network. These dimensions, I will argue, make up any form of network (directed, or undirected, and in any field of study, and within any given context). I will argue that there exist five fundamental dimensions of a graph's canonical structure, and argue how and why the currently known graph-theoretic measures of structure are associated with these dimensions. I will offer clear and unambiguous definitions of these dimensions as well.

After, I will offer a philosophical discussion on risk measurement. These arguments will cement the foundation for my arguments as to why I leverage the measurement framework in this dissertation. Indeed, there has been a plethora of methodologies proposed to measure and classify "risk". I argue in the discussion that the primary reason that this is the case is due to an inherently philosophical disagreement between the varying perspectives in scientific philosophy. I last argue that the disagreements and varying perspectives can be synthesized into a singular framework of which is all encompassing of these differing perspectives, namely, the Bayesian Philosophy.

After this discussion, I will argue that the extant literature has conflated two important and naturally distinct constructs: risk and risk propagation. I will motivate a proper definitional line of demarcation between these two constructs, and argue that it is important that researchers begin to take into consideration the differences between these. Propagation is a relatively new area of study in the supply chain literature. Given that it is still in it's infancy, it would be worth while to offer a formalized definition that distinguishes these two constructs from each other, despite them being related.

One of the major issues in measuring risk propagation is that many presume risk can only flow within a few well-defined paths throughout the network. After my discussion on the line of demarcation of the aforementioned constructs, I will argue that the aforementioned assumption, of which most deterministic and even more recent Bayesian Network models rest upon for their analysis and measurement, is inherently flawed within the study of supply network risk propagation. Indeed, many have applied the

deterministic model of disruption propagation to risk propagation. I will argue that these two constructs, while related, are yet distinct. Furthermore, I argue that there is a major issue with applying deterministic-based disruption propagation models to study risk propagation in supply networks.

Thereafter, I will argue that many of these deterministic models miss a phenomena that I have defined as the "Invisible Path of Causality", something that is inherent within a supply network due to it's various dimensional contextual levels of analysis. I will then argue that the only way to account for these paths of which the prior literature has failed to account for is via the proper application of the Bayesian Philosophy and the various network models within it. I will illustrate, however, that many of whom have applied the Bayesian Network approach have done so inappropriately, and often miss some of the most useful aspects of the theoretical framework (namely, probabilistic reasoning).

After presenting my arguments as to why current deterministic models are insufficient to measure risk propagation, and why current Bayesian Network models have fallen short in proper risk propagation measurement, I will offer a discussion on what we mean by "risk propagation" within a supply network. I will carefully define three constructs of risk propagation. There is no doubt in this author's mind that more exist. As a foundation for future research, however, these three basic interpretations of risk propagation are of the utmost importance to currently focus our attention on, as I will argue they capture most of what many believe "propagation" comprises of.

The primary issue with the topic at hand is that it is inherently mathematical, however, I am going to characterize them from a conceptual perspective. In light of this, I feel the need to justify the approach of conceptual development rather than provide formal mathematical proof. Let me be clear. The theory I have built below is *conceptual*, about something that is mathematical. Some may feel that I should be providing a hard mathematical proof to either motivate or justify my hypotheses. I will argue that this is not necessary, since I am not attempting to understand relationships between what I will identify as "observables", but rather, between the "constructs", which no mathematical equation can fully explain. Hence, I find that there is a need for

me to justify my method of argumentation, so I can proceed with the formal arguments on a solid philosophical footing to avoid the criticisms I would expect from some that tend to view these constructs within a more mathematical frame of reference.

Last, in this chapter, I will develop a collection of hypotheses that seek to answer the primary research questions at hand. I will justify each hypothesis resting on work within the previous literature. Conducting a full-scale mathematical proof for each of these hypotheses would be difficult, if not meaningless, given the conceptual nature of the material at hand. Given that we have conceptual constructs of structure and risk, I feel that the most appropriate way to justify these hypotheses is through a combination of using what has been found in the extant literature in regards to relationships between structure and risk. In this dissertation, I will opt out of mathematical derivations. While they may help to provide theoretical strength between the constructs, I argue that they are not necessary for the topic at hand.

3.1 Supply Chains and Network Structure

3.1.1 The Level-of-Analysis for Networks: A Contextual Perspective

In the literature review, I stated that Peck had organized the construct of a *supply chain* across multiple levels of analysis. These levels were inherently contextual, that is, based on the descriptions of the nodes and edges within what one means when they refer to "the supply network". As mentioned earlier, the precise definition of the *supply chain* is fairly ambiguous, and is entirely dependent on how one characterizes the nodes and edges within the network. (Carter et al., 2015)'s framework avoided such issues by merely using the term *agent*. Their framework for the *Theory of Supply Chain* was indeed general enough to encompass nearly all research that had been conducted within supply chain management (Carter et al., 2015).

However, I argue that when a researcher (and equally important, a practitioner) within the supply network literature (and in practice) seeks to understand some new relationship between constructs, or when they introduce new constructs for that matter, that they should carefully argue as to which level (or levels) of analysis does (do) their

theory apply? Careful definition of the nodes and links within the network from a supply chain contextual perspective is important, as it removes the ambiguities that seem to exist within the extant literature. For example, some within operations research may construct vehicle routing problems within a "supply network". The first fundamental question before taking on such an endeavor would be: "what is the definition of the node, and what is the definition of the link?".

This may seem at first like an trifling question, as the typical answer would be "the firm" and "material flow", respectively. However, even the definitions of "the firm" and "the material flow" are ambiguous without being more specific. For example, there is little doubt that a distribution network comprising of warehouses and manufacturing plants is often considered a "supply network" in vehicle routing models (Melo, Nickel, & Saldanha-Da-Gama, 2009). However, "the firms" would be physical locations, most likely with many locations under the ownership of one firm, and others under the ownership of another firm. If we look towards more of the strategic literature such as joint-venture formation within supply networks, then "the firm" is not a physical location, but rather the entity that comprises of all locations, people, and stakeholders as well as assets, where a single node would represent all of these, and a link would represent a form of a buyer-supplier relationship (either via contract agreement, physical material flow, etc)(Carnovale & Yeniyurt, 2014). That is, the former (the distribution network) is inherently *embedded* into the latter (the organizational network).

Hence, by the above illustration, we can see some potential issues that may arise when theory is constructed absent of a clear indication of the specific supply chain level of analysis. The natural question that arises, however, is: if the level of analysis within a supply chain is important to consider, then what are the levels of analysis in the first place?

Peck offers a brilliant taxonomy of supply network level of analysis, wherein she argued that the supply chain has 4 distinct levels of analysis (Peck, 2005). I will review each of them now. I will argue after, however, that her model is incomplete, and requires revision. Namely, I will argue that there exists one additional level of analysis below her lowest proposed level, and two higher levels of analysis above her highest proposed

level.

On Peck's lowest level of analysis, of which she refers to as the "value stream/product or process" level, nodes represent organizations and links represent more of a "logistics pipeline", where the product "flows" from organization to organization, each one adding value in the process. She argues that both material and information flows through on this dimension of the network. The next dimensional level that Peck describes is that of asset and infrastructure dependencies. On this level, nodes represent fixed assets such as facilities, physical locations, or components of an IT infrastructure such as computer hardware. The "links" on this level represent the connections between these assets, that is, the modes of transportation such as communication, railways, roads, waterways, etc(Peck, 2005).

The next dimensional level up, Peck argues, represents an organizational network. She argues that nodes represent organizations and edges represent dependencies between those organizations such as trading relationships and power dependencies. She argues that this is the "corporate risk management, business strategy and microeconomics" level of analysis. Last, Peck describes the highest level of analysis, which is the "environment" level of analysis. She argues that this is the "wider macroeconomic and natural environment within which organizations do business, assets and infrastructure are positioned and value streams flow" (Peck, 2005, pg. 223).

Unfortunately, Peck does not offer a definition between nodes and links at this level of analysis. In fact, in one of her diagrams demonstrating this multi-dimensional system, she writes "???" at this level(Peck, 2005). For the sake of completeness, I will argue that this level of analysis should be split into two additional levels of analysis. In addition, I argue that this taxonomy of supply chain level of analysis is incomplete. It fails to account for the interactions between firms and consumers, as well as in between consumers. This is especially important to consider, at least from a theoretical perspective, given that social networks now dominate a communication connection between firms and consumers (as well as between consumers).

First, directly above the organizational-network level of analysis should rest an industry-level of analysis. Here, nodes represent individual industries or government

entities and links represent the dependencies of common forms of resources that these industries require to produce value for products that are commonplace within those industries (and, links to the government nodes when there is significant regulatory interaction in those respective industries). Second, at the highest level of analysis, I argue that nodes should represent countries and links between the nodes represent trade agreements and international product flow between the countries.

Last, I argue that there should exist, at the very lowest level of analysis, a "people supply network". The nodes in this network would represent a consumer (where a consumer is defined as an individual of whom has a need for the product, but does not necessarily know about or possess the product), a pseudo-consumer (defined as an individual of whom knows about the product and does not have a need for the product, but has a direct social connection to a consumer or a stakeholder), a stakeholder of whom either directly or indirectly is involved with the product (an employee, an owner, a representative, an investor, etc) and last a government official of whom regulates the product. The "connections" between nodes in this network would represent communication, physical material handling, or financial handling with respect to the product (either its existence, its purpose, its delivery, its manufacturing, its design, etc) between two individuals.

I anticipate much criticism to this conceptual addition to the dimensions of a supply network due to practical concerns. However, I argue that such an addition is necessary to justify certain dynamics between supply chain constructs of which are studied at the higher and more practical levels of analysis. Furthermore I later leverage this newer dimension in order to justify some theoretical aspects of risk propagation. As Friedman states, "the relevant question to ask about the assumptions of a theory is not whether they are descriptively realistic, for they never are, but whether they are sufficiently good approximations for the purpose in hand. "(Friedman, 1953, pg. 14). From a practical perspective, of course it would be impossible to study these networks in their entirety (as it is much greater in size than the organizational-level supply network).

However, certain dynamics between consumers and firms, and among consumers themselves, seem to lack proper justification within a supply chain theory of which

is missing one of the most important actors which serve as it's mere justification for existence: the consumer. The consumer is what drives this existence. Supply chains exist to provide value contribution to a product. But a product can only exist if there is a need from a market, and of course, the market itself comprises of consumers. This is often an overlooked point of contention when academics have attempted to place supply chain on a theoretical basis. While it is true that we had strayed away from understanding *what* we are managing (Carter et al., 2015), we also, in my view, have strayed from *for whom are we managing*.

Hence, this amended taxonomy of Peck's "supply chain" allows the researcher to direct their focus of research towards a specific level of analysis of the supply chain itself. Taking this conceptualization of the supply chain together with Carter's Theory of the Supply Chain (Carter et al., 2015) sets a foundation upon which research in Supply Chain Management can be fruitfully developed while at the same time remaining theoretically well-grounded. It should be noted, however, that despite a supply chain having various dimensions of analysis, a single level is still considered to be a complex adaptive system. Only the definitions (within (Carter et al., 2015)'s theory) of "agent" and "behavior" change when a shift in the level of analysis is made. All else should remain equal. Therefore, given the prior arguments and development above, I formalize the following:

Definition 3.1.1. The *Supply Chain Level of Analysis Taxonomy* is a collection of levels of analysis that describe unambiguously the nodes and links within a supply network. The levels are as follows, from smallest to largest:

1. **The People Supply Network**- Nodes are consumers, pseudo-consumers, stakeholders, or government officials. Edges represent material, information or financial flows between individuals of which are directly related to the product for which the network describes.

2. **The Value-Chain Supply Network** - Nodes represent organizations (more as a proxy to a specific process) that have a specific value-adding process in regards

to the product. Edges represent the same three dimensions, namely material, information and financial flow. This dimension describes more of the inter-relations and integrations between the various value adding processes for the specific product.

3. **The Infrastructure Supply Network** - Nodes represent physical locations and entities such as manufacturing plants, warehouses, computers, etc. Edges represent the transportation mechanism for material, information or money, in relationship to the specific product.

4. **The Organizational Supply Network** - Nodes represent organizations. Edges represent strategic relationships such as strategic alliances, buyer-supplier relationships, and power dynamics, in relationship to the specific product.

5. **The Industry Supply Network** - Nodes represent industries and government institutions. Edges represent the same three flows as before in regards to a type of product (or more generally a need for the product).

6. **The Global Supply Network** - Nodes represent countries and edges represent trading agreements between countries, again, specific to a type of product (or more generally a need for the product).

3.1.2 The Level-of-Analysis for Networks: A Non-Contextual Perspective

Not only is the level of analysis from a supply chain theory perspective important to emphasize during the theory development of any new conceptual addition to nomological nets, but so too is the level of analysis from a canonical structural perspective. As mentioned earlier, supply chains are inherently networks. While I have argued that it is important to provide a clear contextual level of analysis whilst developing additions to supply chain theory, I also will argue that providing the specific network level-of-analysis is equally important. At any of the aforementioned levels of analysis, there exists the network. These level of analyses only speak towards the definitions of the

nodes and the arc. They fail to provide the focal point of *which* nodes and arcs are under study.

Is a single node under study? An arc? A group of nodes? As my prior literature review has revealed, a few levels of analysis from a pure canonical network perspective have be leveraged. Researchers have focused their attention on only a handful of these:

1. The Node Level
2. The Link Level (Also known as the dyad)
3. The Triad Level
4. The Ego-Network Level
5. The Network (or Global) Level

I argue that this taxonomy of structural-based levels of analysis is also incomplete. There clearly is no gap between node and dyad, dyad and triad, triad and ego. However, there is a gap between ego and global. I argue that all of these levels of analysis can be easily summarized through the construct of what I define as the k-ego network. I define this level of analysis recursively for ease of exposition:

Definition 3.1.2. The 0-Ego-Network of a node n or an edge e is the node or edge itself.

Definition 3.1.3. Let n be an element in a graph (either a node or an edge). Let E_{k-1} be the group of nodes in the $(k-1)$ Ego-Network. Then the k-ego network of an element n is the induced subgraph of the union of nodes in E_{k-1} along with the nodes in the respective 1-Ego-Networks of all nodes in E_{k-1}. That is, the k-Ego-Network of an element n is the induced subgraph of all nodes in the respective 1-Ego-Networks of nodes in the $(k-1)$-Ego-Network of element n. If the element n is a node, then it's 1-Ego-Network is simply the induced subgraph of all nodes it directly connects to. If n is an edge, than it's 1-Ego-Network is the union of the 1-Ego-Networks of the nodes of which the node is incident to.

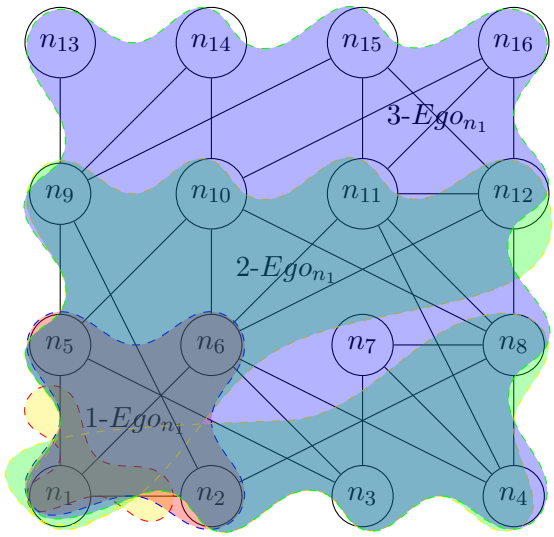


Figure 3.1: The 0-Ego-Network (in yellow area), 1-Ego-Network (in red area), 2-Ego-Network (in green area), and 3-Ego-Network (in blue area) of node n_1.

To illustrate this definition, I direct the reader to Figures 3.1 and 3.2. In both of these illustrations the networks are exactly the same. However, the point of reference is different. In Figure 3.1, we can see that the point of references is node n_1. In this example, the 0-Ego-Network would just simply be the induced subgraph on the collection of nodes $E_0 = \{n_1\}$. Moving one level up, we see that the 1-Ego-Network would be n_1 together with all the neighbors of n_1. This would be the collection $E_1 = \{n_1, n_2, n_5, n_6\}$. Next, the 2-Ego-Network would be all the nodes in the 1-Ego-Network together with the neighbors of all the nodes in the 1-Ego-Network. We can see that these neighbors are $n_9, n_{10}, n_{11}, n_{12}, n_8, n_4$, and n_3. Therefore, the nodes in the 2-Ego-Network would be $E_2 = \{n_1, n_2, n_3, n_4, n_5, n_6, n_8, n_9, n_{10}, n_{11}, n_{12}\}$. Last we have the nodes in the 3-Ego-Network, which again would comprise of the nodes in the E_2 together with the neighbors of all nodes in E_2. These neighbors are $n_{13}, n_{14}, n_{15}, n_{16}$, and n_7. Hence, the 3-Ego-Network would be $E_3 = \{n_1, n_2, n_3, n_4, n_5, n_6, n_7, n_8, n_9, n_{10}, n_{11}, n_{12}, n_{13}, n_{14}, n_{15}, n_{16}\}$, or, in other words, the entire network itself.

It is wise to notice that these levels of analysis are with respect to the node n_1. These levels of analysis change when the perspective from which analysis is conducted changes. For example, Figure 3.2 illustrates that same exact network, only with the Ego-Network analysis conducted from the perspective of node n_7. Following a similar

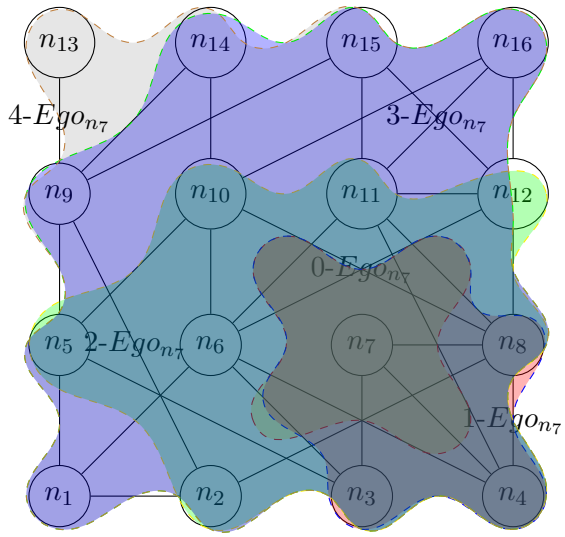


Figure 3.2: The 0-Ego-Network (in yellow area), 1-Ego-Network (in red area), 2-Ego-Network (in green area), 3-Ego-Network (in blue area), and the 4-Ego-Network (in gray area) of node n_7.

breakdown as earlier, we notice that the nodes in the 0-Ego-Network is the set $\{n_7\}$. When we move to the 1-Ego-Network, the nodes are $E_1 = \{n_7, n_8, n_3, n_4\}$.

By the same logic posed earlier, the 2-Ego-Network would comprise of the nodes $E_2 = \{n_7, n_8, n_3, n_4, n_2, n_5, n_6, n_{10}, n_{11}, n_{12}\}$. Next is the 3-Ego-Network, which has the nodes $E_3 = \{n_7, n_8, n_3, n_4, n_2, n_5, n_6, n_{10}, n_{11}, n_{12}, n_1, n_9, n_{14}, n_{15}, n_{16}\}$. Last, of course, is the 4-Ego-Network, which would be equivalent to the entire network: $E_4 = \{n_1, n_2, n_3, n_4, n_5, n_6, n_7, n_8, n_9, n_{10}, n_{11}, n_{12}, n_{13}, n_{14}, n_{15}, n_{16}\}$. Notice how these sub-groups of nodes are different than that of the n_1 perspective. This illustrates that fact that k-Ego-Networks are *relative* to a given node in the graph (and edge, where the "ego network" of the edge would be the union of the Ego-Network of each respective node in the dyadic relationship).

Hence, we can relate the traditional levels of analysis to this more general construct as follows:

1. The Node Level = 0-Ego-Network of a Node

2. The Link Level (Also known as the dyad) = 0-Ego-Network of an Edge

3. The Triad Level $\subseteq$ k-Ego-Network of a Node/Edge

4. The Ego-Network Level = 1-Ego-Network of a Node/Edge

5. The Network (or Global) Level = M-Ego-Network of a Node/Edge, where M is the smallest such number such that all nodes in the network are also in the M-Ego-Network, and vice versa.

While I do contend that this is not an exact equivalence of prior levels of structural analysis, I do argue that prior levels can indeed be related to this new construct as I have illustrated above. This definition of a structural level of analysis allows the researcher of supply networks (and more generally of any field of study that leverages network theory) an easy to use toolbox to (1) clearly specify the level of analysis using only a number of "levels up" from a given location (Node or Edge), (2) encompasses prior levels of analysis that supply network theory has been built on, (3) allows for structural-based measurement of supply network constructs such as "visibility", or a more methodical definition of "the horizon".

3.1.3 Network Characterizations: A Mess of Context

In the literature review, I had reviewed through a few of the basic measures of network structure that past researchers have employed in their research. The issue that I currently identify, however, is that there are so many measures of "structure", and yet, very few attempts to synthesize them into a handful of dimensions. As mentioned, Choi had made a prior attempt at doing this with his three dimensions of "structure". However, the dimensional approach he had taken was myopic with respect to other types of graph-theoretic measures, and he involved conceptual characterizations of supply chain. In addition, his dimensional suggestion was partially rested upon a combination of graph-theoretic measures and context-based measures. While these are useful for explaining strategic dynamics within the supply network, they are practically useless for describing the pure structural components of the supply network. Indeed, Choi simply characterizes "network structure" through the use of only a single dimension, namely "complexity". Yet, there are sub-dimensions of "complexity" that would be fruitful to further explore.

A more thorough examination of the extant literature revealed that no attempts, to the best of this author's knowledge, have been made to reduce these plethora of network measures into only a handful of dimensions. There may indeed be a perfectly acceptable reason for this. From the graph-theoretic perspective, mathematicians often do not rest on dimensionality reduction to characterize mathematical structures. Their characterizations are axiomatic in nature. Furthermore, their characterization of "structure" typically rest on studying a variety of different *morphisms* between structures. That is, they have a tendency to conduct classification or organization of structure based on equivalence classes of mappings between structures, rather than characterizing those structures via inherent properties.

As for the social science perspective on network analysis, I argue that such an attempt for dimensionality reduction and description has not been conducted simply due to two reasons: (1) correlations between the measures are often so high, that they typically take a simple average of the measures (Carter, Ellram, & Tate, 2007) and think of this reduction as simply a means to an end (to reduce correlation between exogenous factors). In addition, (2) social scientists are more interested in understanding how an environment evolves, rather than describing the static characteristics of the environment itself. This may seem counter-intuitive, since a description of the environment is necessary in order to understand an "evolution" of the environment itself (and the results of various behaviors from actors within the environment).

However, in my view, social scientists have taken for granted the characterization (or rather, lack thereof) of the environment (i.e. the network) as something which is a fleeting thought, or a trifling matter, and instead take for granted of the fact that the characterizations are easily measured by graph-theoretic mechanics, that they have no logical reason to study the structure itself. To them, it would seem as though it were a pointless endeavor, since structure is only a proxy to describe the interrelations between actors, which is their more primary pursuit of knowledge.

What is interesting is that any research in social network analysis focuses mostly on the contextual behavior of actors within a network, without first clearly defining the structural attributes of the environment in which they operate. It would be analogous

to Issaic Newton trying to understand the relationship between gravity and movement of particles in the universe, without first describing the nature of the universe itself first. Maybe this is due to social scientists holding a more Objectivist view of science towards their primary field of study. Recall that Objectivism does not assume a context, and rather rests on the belief that there exist only a single "nature". However, from the Relativist perspective, context matters (G. B. Kleindorfer, O'Neill, & Ganeshan, 1998).

In any case, I have argued that there exists a lack of research to properly characterize the structure from a more theoretical and conceptual approach of which is context-free. I may be mistaken, since this literature is vast. But upon a very structured and thorough review of the literature, this author has yet to find an attempt to view "network structure" as a synthesized taxonomy of inherent context-free structural dimensions. There is no doubt that I will receive criticism of this argument that I have put forth. Some may contend that such attempts have been made in the social science literature. However, as Burt mentions, the structure of networks have primarily been focused on actors and relationships between those actors. In regards to measuring the "relationship" or more generally the "link" between two actors, Burt argues that measurement can be carried out by understanding the strength of the relationship as well as the level of joint involvement in activities that both parties engage within.(Burt, 1982).

The primary concern that I raise is that terms such as "strength" and "activity" are inherently rested upon a proper context in which the the network model is being leveraged. I furthermore argue that this research has not much been of interest to the social scientist or the mathematician. This is not a task for mathematicians, as such an endeavor is not in their typical repertoire, or primary area of focus. In addition, social scientists seem to have avoided this endeavor as well by simply assuming that the environment has already properly been characterized by mathematicians (which they have not), via the graph-theoretic measures, or, they feel the need to characterize the network with context-dependent constructs and definitions (such as in (Choi & Krause, 2006), for example). In addition, these scientists view these measures as being more reflective of agent behavior or state than they do of network structure (for example, centrality is not viewed as a structural component, but rather as a means for describing

power or influence of an agent) (Burt, 1982; Freeman, 1978). I therefore argue that a revision to structural analysis of networks is needed. One of which is designed from a context-free perspective.

3.1.4 The Fundamental 5-Dimensions: Towards Characterizing Networks

Every graph, no matter the context, can have a structure of which can be characterized by two fundamental dimensions: *what* is the state of being, and *how* can one operate within that state of being (assuming the structure, or "state", cannot be changed). That is, structure can be characterized in regards to both attribute and description of capacitated and potential behavior within the network. We can attempt to describe the structure of the network from the perspective of a static sense of time. That is, assuming that no actors within the network can "move" within a single unit of time, the primary inquiry rests within the canonical structural characteristics. That is, what is the current state of the structure, assuming behavior is held static? This is contrasted with describing structure based on potential behavior within the network itself, without changing any structural characteristic. That is, describing the structure of the network from a "flow" or "potential flow" based perspective. I therefore contend that all networks, regardless of context, can be described structurally from these two viewpoints: what it *is*, independent of the actor's movements, and how the *is* can be leveraged to *move* within the *is*.

Here is an analogy to rest on during this discussion that will make this dimensional characterization much more intuitive to understand. For example, we can describe the "structure" of the transportation network of New York City from one of two perspectives: what the streets and modes of transportation "look like", and, the ability for individuals to "flow" through that structure. From the former perspective, we most likely would describe the "structure" of NYC according to how many subway lines exists, how many roads there are to travel upon, the average length of those nodes, and so on. That is, we would be describing static-based attributes about the NYC transportation system. On the other hand, we can describe the *ability* for one to "move"

within that structure. Such examples may include the number of paths between any two points in the city, the number of circular paths within the roads of the city, the ability to access other areas of the city during the event of a shut down of a road or a subway. That is, the second dimension, or perspective, describes what actors can "do" within the NYC transit system, while the first dimension merely describes the attributes of the system itself in the static sense.

A traveler in NYC might first know their current location. They then may look at a subway map. The subway map would indicate the locations that the subway system can serve. In addition, the traveler may know from this map how "far apart" the locations are from each other. This, of course, would be describing the subway system from a static perspective. The traveler than may look at specific subway lines that run from one location to another. In addition, they may observe the timetables of the subways that travel throughout the network. The notions of "path" and "movement", or "potential movement" as well as "time of travel", again, are describing the structure of the network, but from a dynamic perspective. Dynamic, not in the sense of changing the subway system itself, or the various subway lines, but rather the current ability to travel within the network itself, and to understand the structure of the network separate from one's ability to move within the network.

Generalizing this idea, we can describe the network structure of any graph or network by considering these two primary dimensions. There exist sub-dimensions, of which will be considered themselves the primary dimensions of this dissertation. In the static sense, the constructs of distance and clustering are "static", in that they don't depend on the actor within the network conducting a form of behavior in order to describe it's current state of being. On the other hand, the notions of feedback (the ability to "loop back" to a position in a certain way), connectedness, and accessibility, respectively, describe the structural characteristics of which depend on (or describe) an action that could be taken by an actor within the network itself. Paying no mind to this distinction between "static" and "dynamic" structural attributes of a network, I will argue that a complete network structure can be inherently characterized according

to five fundamental structural dimensions: Distance, Centrality (Accessibility), Clustering, Feedback (or Thickness), and Connectivity. All graph-theoretic measures can be categorized within any 5 of these fundamental dimensions. I will review and justify each one below.

Centrality

Centrality has been discussed at length in a variety of works on graph theory (Borgatti & Everett, 2006; Borgatti & Li, 2009). Typically it describes how "well connected" a node is to other nodes in the network. The construct of "well-connected" can take on multiple interpretations. For example, some interpretations of centrality have been that it is analogous to "power" or "influence" (Carter, Ellram, & Tate, 2007; Burt, 1982) as well as in regards to how other "well-connected" nodes are connected to another "well-connected" node (Eigenvector-Centrality) (Kito et al., 2014). Another interpretation of "centrality" is how many other nodes it is connected to (degree centrality), as well as how accessible other nodes are, with respect to graph distance, a node is to other nodes (closeness centrality) (J.-B. Kim, 2015). The notion of centrality has been described from an "embeddedness" perspective. It has been used to describe the general position of an overall industry relative to other industries (Nuss, Chen, Ohno, & Graedel, 2016) as well as product positioning (Nuss, Graedel, Alonso, & Carroll, 2016).

Centrality has also been viewed from the perspective of being able to "reach", "share", or "connect" with other actors within a network. That is, the perspective taken from this view is that of "accessibility" (Carnovale, Rogers, & Yeniyurt, 2016). Some have also characterized centrality from a "connectedness" perspective. Rather than a measure of "ability to reach", some have seen it as a measure of "ability to remain accessible" (Y. Kim et al., 2015). This, I argue however, should not be considered a proper interpretation of "centrality", but rather that of "connectedness". While related to the notion of "accessibility", connectedness describes the opposite characteristic (ability to remain accessible in the event of a structure change, verses the inherent characteristic of accessibility).

Yet another interpretation is that of "betweenness", which describes the proportion of paths that travel through a given node with respect to the total number of possible shortest paths between two nodes. This is more of a measure of "relative accessibility or "accessibility relative to a fixed location". In light of all these interpretations of "centrality" in the prior literature, the common trend between them all is that of "accessibility". I therefore will define the dimension of centrality as:

Definition 3.1.4. The characteristic of *centrality* of a node or a network is the ability for a node to reach another node, through a form of travel or flow, within the network.

There are a variety of graph-theoretic measures that can be expected to describe this dimension. First and foremost, the construct of centrality is inherently dependent on that of distance. Node's being "accessible" speaks towards the notion of flow or travel. Upon any travel, if we seek to understand how "easy" it is to travel, one aspect in doing so is that of distance. We can henceforth expect any graph-theoretic measure which involves the basic definition of geodesic distance to be categorized in this dimension. This is simply due to the fact that distance is itself inherently describing the construct of "accessibility", which of course the definition of centrality rests upon. Hence, the following metrics can be expected to comprise of the centrality dimension: betweenness centrality, degree centrality, closeness centrality, eigenvector centrality, eccentricity, and average shortest path length.

Connectedness

Connectedness describes the dynamic ability for one node to "reach" another node in the graph through a walk on the edges of the graph(Diestel, 2017). As I have mentioned above in the previous sub-section, this dimension pertains more towards a network's ability to remain "accessible", rather than it's inherent level of accessibility. Typically there are two notions of "connectedness" in traditional graph theory: edge and node connectedness. Both notions describe how many nodes/edges must be removed in order to disconnect the network. In this sense, the connectedness describes the extent

to which a network can remain "connected" or accessible (i.e. any two nodes can reach each other via a path) in the event of a failure.

This notion has been used to model resiliency of a supply network (Y. Kim et al., 2015). However, this is an improper definition and incongruent with the extant literature. Resiliency refers to "the ability of a system to return to its original state after disruptions" (Y. Wang & Xiao, 2016; Ivanov, Dolgui, & Sokolov, 2017; Blackhurst, Dunn, & Craighead, 2011), rather than it's ability to withstand such a disruption. The construct of "ability to remain connected" or "ability to withstand" is known in the network and supply chain literature as *robustness* (Nair & Vidal, 2011). In this sense, it is a deterministic metric, as it speaks towards the "worst-case-scenario" of which is reflective of the network structure, rather than incorporating randomness. The same type of measures can be designed for collections of nodes like cliques or ego-networks. That is, the scholar of graphs can look past a network's ability to "remain accessible" via edges and nodes and more outwards towards groups of nodes and edges (hence reflecting group-based robustness). Any graph-theoretic measure that relies on the "removal" of a node, edge or combination thereof would capture this underlying construct of "ability to remain connected". I hence expect edge-connectivity, node-connectivity, ego-network connectivity, clique-based connectivity, and coreness measures to properly describe this dimension.

Edge connectivity is defined as the number of edges that must be removed to disconnect the graph. Node-connectivity is defined as the number of nodes required to remove in order to disconnect a graph. When a node is removed, all of it's corresponding edges are also implicitly removed. The k-Ego-Network connectivity, a measure that I have defined in this dissertation, is the number of levels that one must "go up" in k in order for the collection of nodes to disconnect the network.

Coreness of a node is defined based on the k-core of a graph. According to this definition, "informally, a k-core is obtained by recursively removing all nodes of degree smaller than k, until the degree of all remaining vertices is larger than or equal to k" (Montresor, De Pellegrini, & Miorandi, 2013). The coreness of a vertex is then defined as the maximum k such that it lies within the k-core of the graph, but not within the

$(k+1)$-core of the graph. This inherently rests on the act of a "removal", and by my definition, would therefore characterize "connectedness".

Hence, heretofore, I define the construct of connectedness as:

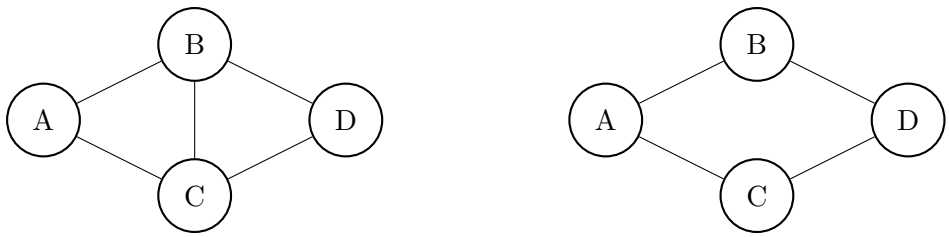
Definition 3.1.5. The characteristic *connectedness* is the ability of a network or an element of a network to remain accessible after the removal or blockage of elements within the network.

Clustering

Yet another dimension of a graph is it's tendency to "cluster" with other nodes in the graph. Indeed, this dimension has been defined as "the degree to which the network is formed of tightly interconnected cliques"(Ahuja et al., 2012, pg. 437). There can be many operationalizations of the notion of a "cluster". Clustering can be understood not from the lens of degree, or "closeness", as intuition might suggest, but rather, of "cliqueness". It describes the extent to which nodes are "close" to each other, but not necessarily in the sense of distance, but rather that of "grouping" or "completeness".

Clustering and centrality are indeed related to each other (Schilling & Phelps, 2007). Upon initial inspection of our notion of "clustering", one may be inclined to define such a construct as nothing more than either an equivalence or a sub-dimension of centrality. However, many of the metrics used to measure the extent of "grouping" or "cliqueness" is not based on distance, but rather that of counting. One of the most popular measures of clustering is of course the *clustering coefficient*. But let us observe it's calculation. It is a ratio of the number of actual triads a node has in it's ego-network divided by the number of possible triads with it's neighbors for the node in the network(Strogatz, 2001). This calculation is conducted naturally by counting.

Another measure of "cluster" is that of "cliqueness". A clique in a network is collection of nodes such that the induced subgraph is itself a complete graph (Burt, 1982). For example, if we have a network with nodes A, B, C, and D, and A is connected to B and C, B is connected to A, C and D, C is connected to A, B and



(a) An example of clustering. We can see that A,B, and C form a clique. We also see that B,C, and D also form a clique.

(b) The density is 0.67 while it's clustering co-efficient is 0.33

Figure 3.3: Caption place holder

D, and D is connected to B, and C, then we can see that the nodes A, B and C form a clique of size three, since the induced sub-graph of A, B, and C form a complete graph (all nodes have an edge between every other node). This is illustrated in Figure 3.3a. The maximum sized clique is therefore found through a counting process. The same holds true for the number of cliques of size k in a network. Another measure that would characterize cliques would be density. This is not necessarily a one to one relationship, however, since high density graphs do not necessarily predict the level of "grouping". However, if there exist a high number or high degree of grouping, then we should expect the densities of graphs to be much higher. For example, it is possible to have a graph that has high density, but low levels of grouping. On the other hand, high levels of grouping may result in higher density levels,although this is not necessarily always true. If we go back to the aforementioned example with the four nodes, with the connections, however, of A to B, A to C, B to D, C to D, then the density is 0.67 (fairly high), while it's average clustering coefficient is 0.33 (fairly low). We can see this illustrated in Figure 3.3b

While the relationship between density and clustering is not very clear, we can contend that density and clustering are at least related in a proportional sense. Hence, this is enough reason to suggest that density may, in combination with other metrics, be reflective of the degree of clustering. Density, of course, is again dependent on a counting procedure in order to obtain its value. Therefore, the notion of counting arcs or sub-structures is representative of the construct of clustering. Variables that would involve this process of counting and of which could can be expected to represent the

extent of clustering within a network would be the clustering coefficient, the number of triads and structural holes, the density, and clique number. The common characteristic in all these interpretations of clustering is the notion of the counting for the means to an end, in order to determine some sense of "completeness". Therefore, I define the construct of "clustering" as :

Definition 3.1.6. The characteristic of *clustering* of a network or a node within a network is the extent to which the network or node exhibits membership within complete groups. That is, it is a measure of "completeness" in the graph-theoretic sense.

Feedback/Thickness

Graphs often have what are known as *cycles* in them. A cycle is a sequence of nodes that one must take from a starting node in such a way that no node or edge is visited twice. The famous "traveling salesman problem" is an example of a problem that involves a cycle in a graph (namely, the "best" cycle). Cycles can explain a graph's "ability to flow back", more specifically, the ability of an actor at one node to traverse in a way that allows them to cycle back. The implications for this type of dimension are tremendous for risk and other related areas in supply networks.

Cycle membership of a node, or the number of cycles and length of cycles within the network, help describe the movement ability within the network itself. While centrality measures accessibility, feedback measures a special form of accessibility. It therefore could be argued that feedback is nothing more than a special sub-dimension of centrality, since it speaks towards the network's ability of accessibility. However, what makes this dimension distinct from the centrality construct is it's interaction with the construct of connectedness. Since cycles describe an actors ability to travel back to their starting location in a specific manner, then the cutting of a cycle would portend a possible level of robustness. In addition, feedback can be argued to be related to the construct of clustering. If we have a graph that is highly clustered, of large cluster sizes or of a large number of clusters, then by definition of the cluster, we must have cycles embedded

within these types of cliques. Recall that a clique is an induced subgraph of nodes such that the induced subgraph is a complete graph. A complete graph of nodes of size three or more would indicate the existence of a cycle. In addition, triads are themselves cycles within a graph. Therefore, a more appropriate argument would be that feedback is an interaction between the constructs of centrality, connectedness, and clustering. It describes the ability of an actor within a network to easily transition back to it's starting place within an linear and non-repetitive manner.

If we look at cycles from less of an "ability to move back" perspective and more from a static structure perspective, then it could be argued that cycles explain a graph's "thickness". That is, from a topological perspective, cycles would themselves determine the "structure" classification of a graph(just as the genus of a topological space explains the structure of that space). Observing the construct of feedback from a static perspective would therefore lead us to the belief that feedback measures the "ability to morph and compress". If we were to compress nodes into other nodes in such a way that we are only allowed to fuse two nodes into one such that doing so would preserve all other paths to and from the same node, then we can see that the feedback construct would lead us to an understanding of minimum (girth) and maximum (circumference) thickness.

The primary method of measuring this construct involves both counting and measuring distance. Understanding the number of ways in which a node, or any node, in a network can traverse back to itself entails a sense of understanding the distance in order to do so. This of course involves a counting procedure. However, the number of cycles is only one characteristic of feedback. The length of those cycles also describes the accessibility of a node to traverse back to itself as well as how "thick" the network is. This interpretation of feedback involves the use of geodesic distance of the graph. Hence, we should be able to expect that feedback/thickness can be measured using the girth, circumference, number of cycles, and average cycle size. I therefore and henceforth define the dimension of feedback as:

Definition 3.1.7. The structural characteristic of *feedback* or *thickness* is the ability

of a node to traverse the network in a way that it does not travel back to previously visited locations to itself. It also can be understood to be how thick the underlying network structure is after a series of compression actions.

Distance

The last dimension is that of distance. Distance describes the "width" or the "length" of the graph, without the action of "compression". Lambert and Choi's early conceptualizations of structure of supply networks rested on this type of dimension. It further describes a nodes ability to access another node in the network, as well as generally how far apart nodes are from each other, and the center. Distance could be argued to be a sub-dimension of centrality. There is no doubt that centrality requires the measurement of geodesic distance in order to properly characterize it. Graph distance requires the same. The natural question that comes to mind then, is, what makes the construct of "distance" distinct from "accessibility"?

While centrality rests on the notion of geodesic distance for it's proper calculations, one can actually consider centrality to be a sub-dimension of distance, not the other way around. Distance measures "how far apart" are the nodes, rather than how "accessible" they are. Of course an argument could be made that supports the idea that these two ideas are one in the same. Again, I will illustrate this with the example I had used above.

The notions of distance and accessibility can be appropriately illustrated using the New York City subway system example. If we would like to speak of the "distance" of the overall network itself, we may be inclined to measure how far apart the stops within the subway system are from each other. This again, however, does not speak towards the "accessibility" within the stops in the network. It may be that two stops from each other are "far apart". In other words, at it's best case, we may need to take multiple subway lines to reach another stop from our starting location. The number of stops in between the two is one description of the "distance" of the subway network.

On the other hand, the ability to travel from one stop to another is not reflected in the notion of "distance". It may be there there are over 1000 different paths we can take starting at our location in order to arrive at our destination. It may not be optimal in doing so, but it describes the ability to get to that destination from our starting point. The former idea of "how big" is the subway network, reflects the idea of the "best way" to arrive at point B from point A. The latter idea of centrality reflects the number of ways in which we can arrive at our destination, not necessarily in the "best way" possible (geodesitcally at least).

Therefore, while centrality reflects a network structure's ability of accessibility, albeit, not necessarily the "best" way possible, the distance reflects the graph's ability to reach a location using the "best way possible". Graph distance is therefore measured using a combination of geodesic distance and the min/max function. We therefore should expect the number of edges and nodes, the diameter, the radius, the distance from the center of the graph and eccentricity to properly reflect this graph characteristic. Therefore, I define:

Definition 3.1.8. The graph characteristic of *distance* indicates the "best way possible" to traverse through the graph. It indicates how "spread out" the nodes are from each other.

3.2 Risk Measurement and Risk Propagation: How Risk "Flows"

3.2.1 A Philosophical Discussion on Risk Measurement

"Risk" in its most general sense is still debated as to it's meaning. As mentioned earlier, some view risk as being solely the likelihood of an adverse event manifesting into something nefarious. Others view it as strictly the potential for something to "go wrong". One of the most common approaches, in SCM, is to consider risk as a mathematical combination of three dimensions: likelihood, consequence, and the sequence of events that can lead to the consequences (sometimes referred to as detection). That is, some

view it as a mathematical product of likelihood, severity, and detection.

While many attempt to place risk within a quantitative framework, there are deep-rooted philosophical problems in doing so that are often overlooked. In fact, I argue that no consensus among risk measure has been reached due to these problems. It could be argued, as I will do so here, that the definition and perception of risk is the result of holding either an Objectivist view or a Relativist view (G. B. Kleindorfer et al., 1998). The Objectivist holds "the basic conviction that there is or must be some permanent, ahistorical matrix or framework to which we can ultimately appeal in determining the nature of rationality, knowledge, truth, reality, goodness, or rightness" (Bernstein, 2011, pg. 8), while the relativist "not only denies the positive claims of the objectivist but ... [holds] the basic conviction that when we turn to the examination of those concepts that philosophers have taken to be the most fundamental-whether it is the concept of rationality, truth, reality, right, the good, or norms- we are forced to recognize that in the final analysis all such concepts must be understood as relative to a specific conceptual scheme, theoretical framework, paradigm, form of life, society, or culture." (Bernstein, 2011).

Risk, as we defined earlier, is a combination of likelihood and severity. The notion of "likelihood" is itself deeply philosophical. From the viewpoint of the aforementioned dichotomy, the Objectivist views "the likelihood" as being somewhat of a constant (not in the sense of change, but rather in the sense of perspective), something that is universal, and that no matter one's observations, the existence of that likelihood is not unique to any particular context or individual. That is, the "likelihood" is part of "nature". From this viewpoint, we hence are studying "what is", rather than "what ought to be" (Keynes et al., 1890). This is not strictly an Objectivist point of view, but rather an economic positive viewpoint of scientific philosophy.

The relativist, however, may view this differently. Since they reject the notion of an all universal logical framework where models and theories can be tested against and within, the idea of having some form of "universal likelihood" is just not possible when such universal environment fails to exist. The relativist views probability as a measure

of truth, but relative to a specific framework or theory. This dichotomy of perspective is personified within the works of Friedman, who comments on the importance to distinguish normative and positive economics. In regards to legislation that many within an economy regard as important to their everyday lives, Friedman notes that "Self-proclaimed "experts" speak with many voices and can hardly all be regarded as disinterested; in any event, on questions that matter so much, "expert" opinion could hardly be accepted solely on faith even if the "experts" were nearly unanimous and clearly disinterested" (Friedman, 1953, pg. 695).

Friedman essentially notes that while normative economics has it's place in regards to finding middle ground in ultimate economic goals across those of whom disagree, positive economics is argued to be no different than any of the other natural sciences. That is, economics follows laws no different than how physics or chemistry does. The point here is that positive economics is more of an Objectivist viewpoint where normative is more of a Relativist view. Friedman's characterization of positive economics being a form of natural science of course does not come without potential criticism. As he states, "of course, the fact that economics deals with the interrelations of human beings, and that the investigator is himself part of the subject matter being investigated in a more intimate sense than in the physical sciences, raises special difficulties in achieving objectivity at the same time that it provides the social scientist with a class of data not available to the physical scientist"(Friedman, 1953).

Furthermore, Friedman argues that theories are not to be judged based on how close their assumptions are to being "realistic", but rather on the ability to empirically validate the predictions that the theory provides. He argues that assumptions cannot be "realistic", but does contend that they can be "approximately" realistic. Indeed, he argues "the relevant question to ask about the assumptions of a theory is not whether they are descriptively realistic, for they never are, but whether they are sufficiently good approximations for the purpose in hand. And this question can be answered only by seeing whether the theory works, which means whether it yields sufficiently accurate predictions. The two supposedly independent tests thus reduce to one test"(Friedman, 1953, pg. 14).

Hence, the Objectivist vs. Relativist dichotomy, along with the Objectivist's sub-theory of Instrumentalism, raises issues in the settling of a singular framework of risk measurement, as it would portend a great debate between the measurement of likelihood in itself. Kleindorfer notes, however, that the Instrumentalist view is not necessarily a pure subset of Objectivism (G. B. Kleindorfer et al., 1998). Rather, the view is an answer to the critics of the Objectivist view. That is, if no theory is to be constructed, then at the least, the Instrumentalist view is a way in which we can organize our observations in order to make predictions.

This dichotomy and the view of Instrumentalism portends a great strife between those who view probability as something of which is inherently subjective and those of whom view it as a measure of the natural world. What is interesting is that probability in itself is a paradoxical answer to a complex problem: that of characterizing uncertainty (how can we characterize something of which we are uncertain?). Indeed, as Popper states, "these [random events] seem to be characterized by a peculiar kind of incalculability which makes one disposed to believeafter many unsuccessful attemptsthat all known rational methods of prediction must fail in their case. We have, as it were, the feeling that not a scientist but only a prophet could predict them. And yet, it is just this incalculability that makes us conclude that the calculus of probability can be applied to these events" (Popper, 1959, pg. 138). He later argues that the paradox can be resolved by taking the subjective view (which is one that he rejects, at least within the scope of statements to be considered scientific or cognitively meaningful, where "subjective" was taken to mean "expectation").

I argue, however, that most managers would prescribe to an Instrumentalist philosophy. As they are very practical in nature, they want a theory that works, and care little as to *why* "it works" (short of the "why" manifesting into some form of cost). A measurement system of risk therefore must be one that appropriately predicts the future manifestation of those risks (i.e. disruptions). However, even this is a claim that is difficult to justify. While managers want something that "works", the criteria upon which something "works" is inherently subjective. Earlier we defined a supply chain disruption as one which results in the supply chain from operating at "normal" levels.

This is arbitrary enough to argue that "normal" is actually a relative construct defined by the actors within the chain itself. What may be "normal" for one firm may not be so for another. So while one firm may experience a deviation from normality, the other firm may not perceive it as such, as their definition of "normal" may be different.

In addition, managers most likely received training within the frequentist philosophy of inferential statistics. Their viewpoint of probability most likely is that it is an objective reality rather than a subjective observation. They view it as something of which needs to be estimated via samples, where the samples are derived from a larger population (in this case, the population would be all possible disruptions that could occur in the past, present and future). If this is true, then any risk measurement framework of which is empirically verified or constructed itself from empirical observation would be circular in nature (managers learn frequentist statistics and hold an objectionist viewpoint, academics design frameworks from these observations, of which managers then subsequently learn).

Hence, given the behavioral characteristics of the inquiry at hand, it is difficult to prescribe a single measurement of "risk" to supply chains. This would explain the vast variation in attempts to measure such a construct in the extant literature. However, this is not reason enough to not attempt to synthesize the measures into a singular framework from wherein these different measures are derived. The differing approaches, I argue, can be synthesized using Garvey's framework, and more generally prescribing the Bayesian philosophy.

As Kleindorfer points out, "the Bayesian approach is concerned principally with dealing with the problem of probabilistic induction (of which the deterministic version is a special case) in a consistent way"(G. B. Kleindorfer et al., 1998, pg. 1095). Risk measurement is inherently a probabilistic inductive activity, despite practitioners and some academics viewing it otherwise. For the manager, probability is seen as a means to an end, that is, as a way to describe the indescribable with the purpose of making a decision in the face of uncertainty.

Given the practical nature of their mindset, it would be difficult to believe that a manager sees probability as a means for reasoning, but rather views it as a purposeful

tool for results (nor would they have any reason to care for it as a means for reasoning). Paradoxically, however, such a view actually transcends their philosophy of pragmatic decision making. When probability is viewed from a subjective lens, the probability eventually converges. Unlike the frequentist, the Bayesianist views empirical observations as knowledge. The knowledge is then taken into consideration to update a previously held belief into a newer belief. The "observation" can be in the form of a single observation (from a single individual) or from a variety of observations (made from multiple individuals). This conceptualization of data is vastly different from that of the frequentist, of whom views empirical observations as being a sample of which originated from a population.

The Bayesian approach to probability and empiricism is apropos for the purpose of estimating risk. The frequentist would argue that the probability of a disruption manifesting can be estimated from prior data on disruptions, surveying variety of viewpoints and summarizing them, or a combination thereof. However, the population of "disruption" is consistently in flux. It changes due to the actors within the network, within the organization, and within the external environment. It would be difficult to model the complexities of this dynamic population of "disruption" from a frequentist viewpoint, and even just as difficult to estimate it.

Hence, I conclude by arguing that the Bayesian philosophy is a good fit for the goal of measuring supply chain risk. The risks inherent within the network are not fixed. It can even be argued that the types of risk are consistently in flux (which would also explain the vast number of classification systems proposed throughout the years). The Bayesian philosophy is easy to consider the "population" as dynamic, as it doesn't view data as being "sampled" from the population, but rather as an inclination of the nature of that population itself. Having an a priori belief in the probabilities of disruption is also natural to how firms operate. Given the diverse definition of "normality" as well as varying perspectives on when the current state of an organization is no longer "normal", the Bayesian approach allows the incorporation of all these philosophical considerations. Therefore, risk measurement in this dissertation will employ the Bayesian approach, as risk assessment and even mitigation can be argued to be forms of probabilistic reasoning.

3.2.2 The Line of Demarcation between Supply Chain Risk and Supply Chain Risk Propagation

Supply chain risk, as I have defined above, is the interaction between the likelihood and the consequence of an event that has the potential to manifest into a supply chain disruption. This definition is very broad and general. It is not specific to a location, type, form, classification, or industry. With that said, the term *propagation* implies that there is some form of *flow* involved with the risk itself. This is distinct, since risk does not necessitate a flow in order to manifest. That is, the action of flowing is a sufficient condition, not a necessary condition of risk.

To illustrate this, let me take from an example. One could decide to play a game of roulette at the casino. The notion of "risk" in this case is again some form of interaction between likelihood and consequence. In this example, and from the perspective of a gambler, risk is charactered by the expected loss of the casino game. That is, what is the average net winnings for the gambler assuming they were to play the game an infinite number of times? The "risk" of loosing or gaining a certain expected quantity of money (which would represent the "consequence" in this example) is inherent only within the action of playing the game. That is, it is self-contained, and its likelihood as well as consequence is not inherently reliant on a sequence of other events in order for the risk to either manifest or "spread". In fact, the risk is only describing a singular and fleeting moment in time. There is no notion of the risk being "caused by" or the risk itself "causing" another event.

This however is contrasted with other forms of risk, that indeed can be characterized by being dependent on a sequence of events, such as a distribution center being out of stock due to sudden shifts in demand patterns, or a machine breaking down and hence leading to a shortfall of order quantities or lead time extensions. The events of "lead-time extension" or "order-quantity shortfall" are inherently dependent on other events that could manifest. The aforementioned events are not "self-contained". They only occur as a *consequence* of another self contained event. For example, an individual does not have a headache for no reason. A headache is inherently a manifestation of

something else that is "wrong" within the body (lack of food, dehydration, something more nefarious).

Therefore, *risk* is not necessarily dependent on a sequence of events, but some risks are. Hence, this line of demarcation between the two forms of events is what would distinguish the definitions between a *supply chain risk* and a *supply chain risk propagation*. The former is all encompassing of the general notion of an event that has consequence and likelihood of which could manifest into a disruption. The latter is a sequence of events that when considered together could lead to a "final" supply chain risk to manifest.

3.2.3 The Problem of Deterministic Propagation

Most of the previous work on risk propagation has assumed that risk "flows" through the network via well defined pathways. While this is true, it is also a bit misleading. Risk certainly "flows" in this regard. However, it is not necessary that a risk be manifested into an occurrence of the event, nor is it necessary to have knowledge (or lack of knowledge) for the risk to "flow".

To personify this point, let me illustrate a simple example. Suppose we have three firms in a supply network: a raw material source, a manufacturer, and a retailer. Suppose that a disruption to production occurs at the raw material source. There is some level of loss at the raw material source, and there is a probability of this event "trickling down" to both the manufacturer and to the retailer. In this case, we know that a risk at the raw materials source has manifested into a disruption. Suppose, however, that the manufacturer keeps enough safety stock on hand of the raw materials to effectively meet their delivery promises to the retailer. Intuitively, we would believe that the disruption has been thwarted and that it would be impossible to trickle further down to the retailer. Indeed, many models of risk assessment rest on this fundamental assumption (Wagner & Neshat, 2012; Nakatani et al., 2018; Zegordi & Davarzani, 2012; Zhang & Lam, 2016; Blackhurst et al., 2018; Liu et al., 2018).

However, this assumption is flawed. There are two ways that a disruption can present itself to the retailer during a similar time as the raw materials. The first is by

mere coincidence. We should not assume that given that a disruption has been mitigated from the upstream of the supply chain that all other disruptions, independent of the primary disruption, somehow magically becomes dissolved. It is entirely plausible to assume that the risk of disruption is still present at a lower node in the network, despite a disruption being thwarted in the upstream. The second way in which a disruption can "trickle down", despite it being "thwarted" above, is due more to a methodical reason as to how risk dependency is modeled.

3.2.4 The Invisible Path of Causality

In the example network, the physical dependencies between the firms are modeled. As such, there is no reason to "connect" the raw materials source to the retailer. They do not conduct business, share information, or maintain any form of a relationship. However, that is not to say that the retailer is ignorant of all that happens upstream. If the retailer obtains the knowledge of the disruption at the raw material source, either dependent or independent of information sharing with the manufacturer, then the retailer may feel that a "disruption" is on it's way.

This in turn may lead the retailer to intentionally force a disruption of operation on itself so as to prevent a possible loss of something of greater value, such as brand value (this strategy can be somewhat thought of as a "vaccination approach" to mitigation; that is, accept some level of known pain now with an intentional self-inflicted disruption rather than wait for an unknown level of pain and uncontrollable disruption later). For example, it may decide to shut it's store down for a day until it gains confirmation from the manufacturer that the disruption has been thwarted and deliveries are on schedule. Another possible scenario is that the market itself could gain information regarding the disruption at the raw materials source, and hence, an abrupt shift in demand patterns within the market occur (hence leading to a disruption at the lowest point in the chain).

It could be argued that such a discrepancy could be accounted for by establishing a link between the raw material's source and the retailer (and between the raw materials source and "the market"). From a methodological perspective, this would indeed solve the problem, and full paths can be considered again from a deterministic perspective.

However, a loss of meaning of the network itself would ensue. If we follow the reasoning posited above, then risk can spread from any node that is in the upstream to any node that is more downstream, as I illustrated earlier. If we seek to solve the problem illustrated above by adding links, then it would be necessary to establish a link from all nodes to all other nodes. That is, every supply network, or more generally, a model of risk within supply networks, would simply be a complete graph (which is a network where every pair of nodes has a link between them). The underlying structure and meaning of the network would dissipate, and become theoretically useless.

In supply chain, we need a sound and logical reason to model a connection between two firms. We cannot simply justify a connection between two firms by claiming that it is possible the one firm is able, or possibly able, to acquire information about the other firm. A link between two firms in any model of a supply network needs to be justified from some form of meaningful relationship between the two firms (be it tactical, operational or strategic). This is a cornerstone of social network theory, of which supply network theory is built atop. For example, if you happened to see Kevin Bacon at a local Starbucks, does this equate to you now having a "tie" with Kevin Bacon? I would speculate in the negative. You may have obtained information on Kevin Bacon's favorite drink to order, but this is not reason alone to establish a link in your own social network to him. Therefore, such a modeling approach is not an appropriate solution to the aforementioned "issue".

This "issue" is inherently a form of bounded rationality of the agent in the supply network (be it a manager or an analyst). Recall, I argued earlier that risk measurement (and more generally, assessment) is inherently subjective. However, despite the bounded rationality of the analyst, this does not negate the existence of a series of complex interactions between two events that may not have any "logical" reason to be connected, but rather only so through means that are ineffable. In the aforementioned example, I mentioned two ways in which disruption can ensue at the retailer. In actuality, there are probably an inconceivable number of "paths" of causality between the raw materials source and the retailer through nodes that were not considered, yet are located within other dimensions of a supply network (i.e. a people-supply network, per se). These

unseen paths through these unseen nodes are what I am defining as the "Invisible Paths of Causality".

Essentially, I am arguing that despite the limitation of the analyst to foresee every possible event and interdependence between the events, there still, none the less, exist the possibility of one known event to "cause" another event, despite the path between them forestalling such a scenario. For the modeler to pursue an interest in identifying these invisible paths would not be worthwhile, practical, or cost-effective. This is why Garvey et. al had argued in favor of constructing a network of risk factors that is inherently derived off the structure of the supply network. The reason rests in Occam's Razor: it is simple to model, and hence, is likely the "correct" way to model it.

Most analytical models of risk strive towards this notion of simplicity. As with any theory of, well, anything, "one of the most common mistakes of beginners is the inclusion of too much detail in their models. The old hands always tell novices to make their models as simple as possible by including only first order effects" (G. B. Kleindorfer et al., 1998, pg. 1092). On the other hand, however, by ignoring these invisible paths of causality within a supply network, we focus directly on the real supply network structure (rather than on a higher-dimensional, more abstract and more difficult to describe supply network), and may miss the potential to model these paths of causality, which may be as detrimental to the firm as well-conceived known paths of causality.

If there exists a model that can accurately predict the occurrence of one event, given knowledge of an indirect event (occurred, or, did not occur), yet with no "logical" path as traditional analysis demands, using only the structure of the supply network as a structural criteria, then we have fulfilled Friedman's conceptualization of a "good" theory. Indeed, a "theory is simpler the less the initial knowledge needed to make a prediction within a given field of phenomena"(Friedman, 1953, pg. 698) Short of Popper's and Friedman's treatment on the simplicity of models and theory (Friedman, 1953; Popper, 1959), I think it stands to reason that attempting to chase after this "Invisible Path of Causality" from a more analytical approach is inherently foolish. Yet, a lack of knowledge of this type of causality path is a major drawback to current models of risk propagation (of which I will review shortly). So then, how do we keep the

simplicity of the risk propagation model, but also try and account for these "Invisible Paths"?

3.2.5 Balancing Model Complexity and Error in Risk Propagation

I say "issue" because it really is not an issue when risk propagation is modeled using Bayesian Networks. Just I have have argued above for their use in risk measurement, a Bayesian Network is useful for risk propagation measurement (and, measurement of risk characterized via propagation). The framework of Bayesian Networks allow for the capturing of these so called "invisible paths of causality" via the conditional probabilities. In the previous example, the probability that a disruption will occur at the manufacturer, given that a disruption has occurred at the raw materials source, may be observed to be, for example, 0.6. As for the retailer, we may find that the probability of a disruption at the retailer, given that a disruption has occurred at the manufacturer, may be 0.2. In the traditional frameworks of risk analysis, this would essentially be enough information to use to prescribe a set of risk scores to the events. However, this is being myopic, given that the model specification is incomplete.

Indeed, what was not specified in the example above was the probability that a disruption occurs at the retailer (and the same for the manufacturer), given that a disruption *did not occur* at the manufacturer. This probability may be, for example, 0.6. This example may seem paradoxical to intuition. However, I have already implicitly argued as to how this may happen. Why is it that the probability of a disruption at the retailer may only be 0.2 if a disruption occurs at the manufacturer, but would be 0.6 when such a disruption does not occur at the manufacturer? One possible explanation could lie within the contract that is between the retailer and manufacturer.

A disruption indeed may occur at the manufacturer, but maybe their contract required the manufacturer hold a certain level of safety stock to be allocated and used for that specific retailer. In this case, even if a disruption were to occur within the manufacturer, they would be able to contain it to within their own firm. What then dictates the probability of a disruption being higher (0.6) if it is known that a disruption did not occur at the manufacturer? This may be due to a disruption occurring at

the raw materials source. As I argued earlier, knowledge can travel to the market, or the retailer obtaining that knowledge can induce a self-inflicted form of a disruption. In such a case of an "invisible path of causality", the Bayesian model would be able to take into account these intricacies without further complicating the model from a structural perspective.

To further elaborate on this notion of "invisible paths of causality", there is yet another portion of this that the extant literature seems to ignore, ironically especially within the literature that leverages Bayesian Networks. One of the key benefits of using Bayesian Networks is due to the ability to reason probabilistically. In the aforementioned example, the retailer may have knowledge of a disruption at the raw material source, but may not know if the disruption "trickled down" to the manufacturer.

The extant literature confounds the notion of "no disruption" and "no knowledge" (Qazi, Quigley, & Dickson, 2015; Qazi, Dickson, et al., 2018; Qazi, Quigley, Dickson, & Gaudenzi, 2015; Qazi, Quigley, Dickson, Gaudenzi, & Ekici, 2015; Qazi, Quigley, Dickson, Gaudenzi, & Önsel, 2015; Qazi, Quigley, & Dickson, 2018; Qazi et al., 2017). These are very distinct from each other. In the more deterministic frameworks of reliability or event-failure modes, it is assumed that each node has one of two states: operating or not operating (in the case of disruption and what some refer to as "risk", it would be "occurred" or "did not occur"). However, in the Bayesian Network framework, if each node is modeled with two states, then in reality, it actually has three states (the event occurred, the event did not occur, no knowledge of the event).

When knowledge of the event is ascertained, it is referred to as *evidence*. Therefore, for the retailer, having information regarding the disruption at the raw material's plant, would not be able to conduct a proper analysis using a more traditional deterministic approach if knowledge is not known about the state of operation at the manufacturer. This is the distinct advantage of leveraging a Bayesian Network. The network allows the evidence of disruption at the raw material source to *flow* to the manufacturer. The belief of a disruption propagating to the retailer can then be *inferred* through a complex algorithm known as *belief propagation*, despite having no knowledge of the current state at the manufacturer.

I will now illustrate the "Path of Invisible Causality" with a small demonstration, and also emphasize via this example as to why Bayesian Networks are the appropriate solution to capturing these inconceivable paths. Let us begin with a simple example, the same one as above. Suppose we have three nodes: a raw materials source, a manufacturer, and last a retailer. There is a link in the supply network from the raw materials to the manufacturer. In addition, there is a link from the manufacturer to the retailer. Let us say, for illustration, that there is a link between the raw material's source and the retailer. There is no logical reason for this connection other than the fact that the retailer has the ability to acquire information regarding the raw material's source, either directly or indirectly. As I have previously argued, adding such a link is not quite in line with traditional theory on SCM. As there is no level of reasonable argumentation as to why these two nodes should be connected, then, from a modeling perspective, they should be disconnected.

However, again, as I argued earlier, if an actor within the retailer gains information from an actor within the raw materials (that is, via the lower dimension of supply chain structure, the "people-supply-chain"), then the raw material disruption would indeed impact decisions made at the retailer, despite no logical reason (at the higher level of analysis) to model a connection between the two. The link, in this example, is justified by a link in the "people-supply-chain", despite such link having no meaning in the "organizational-level". Let's suppose now that a disruption *could* occur at the raw material source. Then both the manufacturer will have a probability of disruption and the retailer will have a probability of disruption (since the raw material and retailer are linked via this lower-dimensional connection). Within the Bayesian Framework of Analysis, we therefore must model the probability of a disruption at the retailer as being dependent on what is known at (1) the raw materials source and (2) the manufacturer. Suppose the conditional probability table comprises of the likelihoods as demonstrated in Table 2.1. Furthermore, suppose the probability of a disruption at a manufacture and the retailer, respectively, are as in the same table. Under this model of risk, the "invisible path of causality" would be the link between the retailer and the raw material source. However, due to the mechanics of Bayesian Networks, we can "fold" the link

	Raw Material = 0		Raw Material = 1	
P(R)	M = 0	M = 1	M=0	M=1
R=0	0.6	0.2	0.1	0.7
R=1	0.4	0.8	0.9	0.3

Table 3.1: Conditional Probability Table of Disruption at Retailer Given Manufacturer and Raw Material

P(RM)	
RM=0	0.4
RM=1	0.6

Table 3.2: Conditional Probability Table of Disruption at Raw Material

between these two into a simple chain from the raw material, to the manufacturer, and to the retailer. That is, we can take this more unrealistic risk analysis (since it is assumed to be unrealistic to model risk between the retailer and the raw material source), and *fold* it into model that "makes sense" from an organizational network perspective. That is, there will be no link between the raw material and retailer, but rather, only indirectly via the manufacturer. Let us observe the mathematics as to how this compression can occur.

Ideally, we would like to find $P(Retailer = 1 \mid Manufacturer = 0)$ and $P(Retailer = 1 \mid Manufacturer = 1)$. Finding such probabilities would result in a complete specification of the more realistic probability model. In the less realistic probability model, however, we have another parent of the Retailer, namely, the raw material source. However, using the law of conditional probability, from Bayesian Analysis, we can find the aforementioned probability by conditioning against the raw material source:

$$\begin{aligned} P(R = 1 \mid M = 0) &= P(R = 1 \mid M = 0 \wedge RM = 0) \cdot P(RM = 0) \\ &\quad + P(R = 1 \mid M = 0 \wedge RM = 1) \cdot P(RM = 1) \\ &= (0.4)(0.4) + (0.9)(0.6) \\ &= 0.7 \end{aligned}$$

and:

$$\begin{aligned} P(R = 1 \mid M = 1) &= P(R = 1 \mid M = 1 \wedge RM = 0) \cdot P(RM = 0) \\ &+ P(R = 1 \mid M = 1 \wedge RM = 1) \cdot P(RM = 1) \\ &= (0.8)(0.4) + (0.3)(0.6) \\ &= 0.5 \end{aligned}$$

Therefore, the less logical network can be compressed in the "more logical" form by the distribution of $P(R = 1|M = 0) = 0.7$ and $P(R = 1|M = 1) = 0.5$. Hence, we can see that the "invisible path of causality", which was modeled above with a link between the raw material's source to the retailer, can be compressed into a form where no such link exists (the more logical form shown in Figure). Furthermore, if these probabilities were estimated from a dataset of observations, then we should see the same probabilities illustrated above (in the two separate models).

Hence, I have illustrated here that if the "invisible path of causality" is modeled as a link between the two nodes in the Bayesian Network, that this link can be "compressed" into a network that is more "logical" according to the level of analysis. Furthermore, I have demonstrated through this analysis that the probability of disruption at a lower node can actually be higher (0.7) when no disruption occurs at the direct parents (the Manufacturer) than the probability of a disruption at the Retailer when a disruption occurs at a Manufacturer (0.5) due to this "unseen" path of causality from the raw material source to the retailer.

If the analyst is presented with proper disruption data, then the estimation of such a Bayesian Model **will** result in similar outcomes. That is, despite all prior comments from the previous literature on the matter, I have illustrated that accounting for "disruption" at lower nodes (such as the retailer) despite no such disruption occurring at the manufacturer, while it occurring at the Raw Materials source, is of the utmost importance due to this "invisible path of causality". Again, this path only becomes visible when the analyst is present with enough data, or chooses to adopt a model of causality via Bayesian Networks ignoring the supply network structure. As I have argued above, adopting such an approach is misleading, and most likely too time consuming. If we

have data regarding disruptions at the lower levels of the supply chain, despite them being "thwarted" at higher level nodes, then we can easily estimate the chances of a disruption, despite knowledge of the disruption being "thwarted" at the middle level.

Previous models of supply chain risk have rested on deterministic models such as Petri-Nets and Event-Failure models. While there is absolutely no doubt these are sufficient for conducting a disruption analysis, I have argued here that they fall short in accounting for these "invisible paths of causality". The modeler of deterministic events is hence left at a fork in the road. Either they ignore the structure of the supply chain network and apply their more traditional methods of fault analysis to a network that is (1) complex and (2) non-nonsensical from a theoretical supply-chain point of view, or, they must abandon the possible link of causality all together, which of course, will inflate their errors in their analysis. This argument alone is sufficient to justify that fault-analysis and reliability-based methodology towards risk propagation analysis in supply networks is a very poor measurement and methodological approach. It is almost quite near literally fitting a square peg into a round hole. Henceforth, I submit here, that for risk propagation analysis, the **only** model that can properly capture the interdependencies of risk is a Bayesian Network model (and, possibly the generalization of such a model, such as Dynamic Bayesian Networks).

3.2.6 Constructs of Supply Chain Risk Propagation

I defined earlier that "risk" can be understood as the *interaction* between likelihood and consequence. I adopted this definition since there are many interpretations of "risk". However, the adoption of this definition leads to some slight construct measurement issues. For one, a manager may view "risk" from a vastly different perspective than a different manager. From a high-level management perspective, "risk" may be viewed as a disruption occurring at the firm level, and affecting not only the firm itself, but the firm's relationships. Strategic Management seeks to address the continued survival of the firm by ensuring that it's performance is operating at an acceptable level to due so. Over the prior few years of theory development in the literature, it has been shown that part of this survival rests on not only one's own firm surviving, but others within

it's supply base (or, more generally, it's ego-network).

Therefore, a corporate-level manager may see "risk" from the lens of not only disruptions occurring within the firm, but also helping to mitigate any of those risks from spreading to lower parts of the supply chain (and upstream members). A production manager, on the other hand, may view risk much differently, and within a much more refined scope. They may view "risk" simply as a disruption that occurs within their own firm, and may not particularly care if that manifestation of the disruption continues to "trickle-down" to the next level, so long as their firm is "spared" of the cost of the disruption. This may seem myopic, but it also may be due simply to the role and level the manager holds for their position within the organization.

The production manager most likely would like to measure risk in terms of the chance of a disruption, and even the consequence of it, only within the firm itself. If an event is "risky", but the consequence to the firm itself is low, then the measurement of risk of which the manager seeks to have reflecting the construct would most likely be solely focused on the consequences to the firm itself. The corporate-level manager, on the other hand, may view "risk" differently. They may view risk from the perspective of a total cumulative effect within the supply chain. They, after all, are forward looking, and seek to ensure that their corporate strategies are effectively carried out with minimal disruption to them.

These may involve the collaboration and coordination between and among multiple firms, in multiple locations, at multiple different points in time. Therefore, unlike the production manager, the corporate manager most likely would like a measure of risk that reflects the cumulative damage to the chain, and not only the consequence of a disruption within their firm. While an event may seem unlikely, or may be low in consequence to the individual firm, where the production manager sees no issue, the corporate level manager sees a problem (the propagated consequences may be more severe than the consequences within the focal firm itself). As such, multiple metrics of risk are necessary for different purposes of measurement, as well as applying to different types of agents acting within the network.

Three fundamental measures that can come to mind when viewing risk through the

lens of a "supply network" are as follows: where else did the risk occur upstream, where else can the risk spread to downstream, and to what extent does the current node in the network contribute to the difference between these two extremes? That is, risk can be viewed from a diagnostic perspective, a causal perspective, or from a "change" perspective. I will now address all of these below.

Propagation Diagnostics: Risk Significance

Upon the observation (or simulated observation) of a disruption, one question that of course would come to mind is: from where did it originate? This of course is the more general process of diagnosing a business problem. Knowing what had caused a particular observation made within the firm would give the power of knowledge to the firm to best adjust any strategy, if necessary, as well fine-tune any controls within their own organization.

The process is analogous to the one that a medical doctor often conducts. When a patient comes in with certain symptoms, ideally, it is the doctor's goal to understand what may have caused the symptoms themselves. Understanding this would lead the doctor to the primary, or handful of primary, reasons for the symptoms. They then can address the primary problem at hand, rather than separately prescribing medications to fix the symptoms instead of the core problem (or problems). As we know, fixing symptoms may lead to temporary relief for "pain", but would be ignoring the longer-term issue, and hence, recurrence of the symptoms should be expected. A supply chain manager must take the same approach.

That is, when a disruption occurs, it is the job of the supply chain manager to understand the underlying reason for the disruption itself. Such an understanding allows the supply chain manager to ascertain further knowledge of the disruption. This knowledge, in turn, can help further understand the potential solution and immediate action to take to either help reduce a further worsening of the disruption or to help outright prevent it from occurring again in the future. In addition, even if this knowledge is not directly usable by the focal firm, it may be used for either strategic reasons or competitive reasons. The firm could be highly cooperative with other firms in it's network. If

such firms form a tight clique, they most likely share information regarding disruptions so as to have the entire clique "optimized", rather than just the firm. On the other hand, if the firm is more opportunistic of disruptions, then such knowledge could help the firm gain a competitive advantage that was not previously realized. Particularly the knowledge of disruption, which could indeed affect changes to pricing or allocation policy.

Therefore, a useful risk metric to have in the manager's tool box would be how "significant" a node is within the supply network with respect to diagnostic-based risk. That is, given information regarding the occurrence of a disruption at the current location of focus, where else, or, more appropriately, how much additional damage should we expect to see within the upstream of the network? The *risk significance* would therefore indicate how much damage was necessary, or expected, in order to "trickle down" to the current focal node in the network. This risk is "significant" in that it indicates a level of "robustness" for the current focal location. Higher levels of this construct would indicate a certain level of significance for the current node. That is, high levels for this node would essentially be "better", since we need to expect a larger amount of risk to manifest in order for it to manifest within our own location. Lower levels of course could be interpreted as being "worse", since fewer disruptions need to occur in order for the disruption to eventually manifest in the current location. Therefore, if we are observing this construct from a robustness perspective, "lower" risk significance is "bad", in that our current node is very sensitive to occurring from only a handful of disruptions. "Higher" risk significance would be "good", in that we need to expect the disruptions to "trickle down" in a much larger extent.

This construct could of course be interpreted from the opposite perspective. If we happen to notice that we have a low "risk significance", then the cost that we must expect upstream needs to be low in order for the cost at the current location to manifest. Depending on the expected consequence to occur in light of a disruption at the current location, then, a "low significance" would indicate a low cost, which means that the node is of little concern. High significance would translate to higher cost. Therefore, if the significance is higher, we should expect higher costs across the upstream members.

Therefore, the interpretation of "risk significance" really depends on how one defines "cost" as well as their behavioral attributes (are they more opportunistic, or are they more cooperative?). Regardless of perspective, the measure simply strives to reflect one property: how much "risk" can be expected "upstream"? When I say "upstream", I am referring to the direct ancestors of the location. This would exclude any damage lost or disruptions observed at children of ancestors of a focal firm that are themselves not ancestors. This definition of risk is refined to the scope of only those nodes in the network that have a one-way directed path of risk propagation. Therefore, I define the following construct of risk:

Definition 3.2.1. The *supply chain risk significance* of a location in a supply network is the possible or potential level of damage that could have occurred at the ancestors of the location upon the observation of a disruption, or no disruption, at the aforementioned location in the supply network

Propagation Causality: Risk Contribution

Some view risk from the lens that it "spreads". That is, not only can risk impact a single location within a supply network, but can also continue to affect others within the downstream, as a result of either mitigating or not mitigating it at a particular location. If a risk were to manifest into a disruption, a question that may be of interest to the practitioner may be "where else can I anticipate the disruption to spread?".

Unlike significance, the *risk contribution* considers the potential *a posteri* impacts of a disruption (while significance considers the *a priori*). From this perspective of "risk", it is viewed more within the control of a location, which differs from the risk significance. That is, we can consider a location to be "risky" if it has a tendency to spread any disruptions that occur at it's location (or, if it has a tendency to thwart disruptions locally, but manages to spread risks of disruption downstream).

In a supply network, the *downstream* comprises of those nodes that are direct descendants of the a focal location. It may not be as linear as we intuitively think it is. The customers may themselves have multiple customers, and so on and so forth. Since these customers are partly dependent on the supply chain activities of the focal

node, it is logical to assume that a risk of disruption "spreading" down stream to these descendant nodes is possible in the face of a disruption.

An analogy to this interpretation is determining the likely symptoms of which may manifest if a patient is known to have a certain sickness or disease. The medical doctor will analyze the patient, and attempt to understand which symptoms are most likely to manifest as a result. This of course is yet another form of business problem solving, primarily within the area of decision analysis. Often, practitioners attempt to understand the potential consequences of their decisions, and from there, make decisions in a way that minimize the negative consequences.

A measure of risk that could prove useful for practitioners is that of how risk spreads downstream. Decision making at a focal node, as argued earlier, may manifest into disruptions at a later point in time. If we understood the total cost of the downstream if a disruption were to occur (or not occur), then we would have a proper measure to use to gauge decision making when facing the uncertainty of a supply network risk spreading to the downstream of the network.

Hence, I define the following:

Definition 3.2.2. The *supply chain risk contribution* of a location in a supply network is the possible or potential level of damage that could occur at the descendants of the location upon the observation of a disruption, or no disruption, at the aforementioned location in the supply network

Propagation Change: Risk Velocity

While cost may be of a concern to practitioners, so too are metrics that indicate the "speed" or rate of change of those costs with respect to "flowing" through a particular location. Risk in a supply network is no different than this conceptualization of "speed of spread". While the cost at one location may be low in the upstream, it may be higher in the downstream as a result of "flowing" through that location versus an alternative location.

Understanding how risk "changes" as it "flows" through different locations would

help a manager identify potential "enablers" versus "depressors" of risk as it flows throughout the network. Such a metric would aid managers in understanding which mitigation strategies have worked in preventing the flow of risk and which ones further accelerate risk itself.

The word "velocity" comes from physics, which is often defined as a "change in position within a given quantity of time". If we ignore the notion of time for the moment, and instead look at change with respect to "change in location", then we can relate the construct of "velocity" to that of risk. In addition, future measures can be constructed such as acceleration, which may help further aid the practitioner in identifying truly "risky" nodes when "risk" is viewed more from a perspective of "contagion" or "rate of spread" than it is from a cumulative loss perspective.

Hence, I define the following:

Definition 3.2.3. The *supply chain risk velocity* of a location in a supply network is the rate of change in the possible or potential level of damage within the ancestors and descendants of the focal location, with respect to the change in moving from the upstream to the downstream, upon the observation of a disruption, or no disruption, at the aforementioned location in the supply network

3.2.7 Methods of Operationalization for Risk Propagation Constructs

Earlier in the dissertation, I reviewed through a variety of prior methods that have been suggested in the extant literature in order to measure risk. There have been many approaches that have ranged from using subjective-scaled responses, AHP, DEA, ISM, Fault-Analysis, Petri-Nets, Graph-Theoretic Based Measures (Connectivity), Value-at-Risk (VaR), worst-case-scenario, and expectation-based methods, as well as other computations based on random variables and the use of their respective probability distributions.

As I have argued in the previous sections, the Bayesian Network approach allows for the research of supply chain risk to easily measure and account for a variety of different

forms and types of risk. The conceptualizations of supply chain risk propagation outlined above can easily be operationalized by leveraging the Bayesian Network approach (as well as other approaches). The first step in doing so is of course to define what one means by "risk" when one is conducting research (or attempting to measure it).

As I have reviewed earlier, the notion of "risk" is inherently subjective. Often times, it's interpretation depends on how it is being involved within the particular system of study or operation. Above I outlined three fundamental constructs of supply network risk propagation. I intentionally defined them conceptually, albeit not arbitrarily, since, to the best of this author's knowledge, no prior attempt at conceptually defining "supply chain risk propagation" has been conducted. The definitions I have set forth in the previous section can be used as the baseline for understanding risk from a propagation perspective within the supply network.

When it comes to research, however, we must find ways to operationalize these constructs. Recall that I have defined supply chain risk as an interaction between the consequence and the likelihood of the consequence occurring. This notion of "interaction" between these two ideas is implicit within the definitions (via the use of the phrase "potential consequence"). However, when conducting research, the word "interaction" could be interpreted a variety of ways, as I had illustrated earlier in this dissertation.

Therefore, proper operationalizations of the risk propagation constructs would require the researcher of supply network risk propagation to first establish what they mean by "interaction". This could mean ignoring the "positive" outcomes of an event and simply multiplying the likelihood by the consequence. This could also mean "expectation", where the traditional mathematical expectation operator is used on the random variable, or random variables, that describe the various risks. In addition, one could also take the perspective of interaction meaning a "certain percentage of", just as in many Value-At-Risk models. A variance or more general moment's perspective can also be taken.

Therefore, operationalizing these constructs involves first (1) identifying the risks themselves and establishing how they are to be modeled (as scores, random variables,

etc), next (2) by establishing how the likelihood, consequence, or both are operationalized (i.e. what does the researcher mean by "likelihood", and further, what do they mean by "consequence" or "cost"), and last (3) establishing the nature of the "interaction" between the likelihood and consequence (none, if only one of these are chosen to represent "risk"). Again, the "interaction" is what defines the risk constructs above. Hence, in order for the risk constructs to be properly operationalized, the notion of "interaction" must also, and it must be clear from the research of risk propagation what specific type of interaction between the two is being used to model "risk".

In the research methods chapter of this dissertation, I will explain the specific forms of operationalizations I use in this dissertation (mostly that of the expectation operator). I want to make it clear that such an operationalization should not result in a loss of generality for describing how the constructs of, well, anything, and risk propagation, are related. Hence, I argue here, that when researchers seek to operationalize these constructs of risk propagation, they must offer clear and unambiguous operationalizations or interpretations of "interaction".

3.3 Hypothesis Development

3.3.1 Conceptual Development

2-Ego Network Risk

2-Ego Network Risk is dependent on the risk of the nodes and arcs within the 2-Ego Network. It is, conceptually, a "collective" risk of the group. As the one of the primary topics of this dissertation, I seek to examine if the overall network structure of the 2-Ego Network would affect the the overall risk of the 2-Ego Network. As I have defined above, every network structure has the five dimensions: Centrality, Connectivity, Feedback, Distance, and Clustering. To the best of the this author's knowledge, no prior work has been conducted that observes any potential relationship between these constructs (relative to a 2-Ego Network) and the construct of supply chain risk propagation. As a reminder to the reader, I will be using the phrases "supply chain risk" and "supply chain risk propagation" interchangeably. Please keep in mind, these are technically

distinct. For the sake of brevity, however, I will drop "propagation", and the reader should be aware that whenever I used "supply chain risk", I am referring to "supply chain risk propagation".

We can speculate, that we should not expect the structural design of 2-Ego-Networks to affect the 2-Ego-Network risk in the same manner as that of 1-Ego-Networks affecting 1-Ego-Network risk, unless the two structures are exactly the same. First, 2-Ego-Networks will be at an absolute minimum the same as the 1-Ego-Network. In this instance, structure would impact risk in the same exact manner (since the networks would be the same). However, it can be reasoned that this trivial case is unlikely to occur in realistic supply networks. For this scenario to be actualized, all firms in a focal firm's ego-network would themselves need to have the same exact ego-network as the focal-firm.

Previous literature has suggested that firms can have as much, at the low end, as 200 suppliers (Kito et al., 2014; Carnovale & Yeniyurt, 2015). Therefore, the probability that a firm's ego-network would be *isomorphic* onto it's 2-ego-network would infinitesimally small. Given this, we know that it's more likely for a firm's 2-ego-network to be different in structure than it's ego-network. By definition, however, the firm's ego-network would be a sub-graph in it's 2-ego-network. Since we have argued that it is unlikely for a firm's 2-ego-network to be isomorphic to it's ego network, we therefore are left to conclude that a firm's 2-ego-network will most likely be larger than it's ego-network.

Since a firm's 2-ego-network will most likely include nodes not in it's ego-network, then we should expect the distance from these nodes to the focal firm (and other nodes in the ego network) to be larger (by definition). Therefore, we should expect

Hypothesis 1. *A focal firm's 2-Ego-Network Distance, ceteris paribus, is greater than or equal to it's 1-Ego-Network Distance.*

Since the 2-ego-network of the focal firm will include nodes that are not in the ego-network, the nodes in the 2-Ego-Network have greater accessibility to other nodes within the network. While the distances in the ego-network will generally be higher,

there will indeed be more connections within the 2-Ego-Network, since there are more nodes within the 2-Ego-Network. Recall that accessibility is the construct of the "ease of ability to reach" another node. From the network perspective, this entails observing the general accessibility starting from any node in the network and "reaching" another node. We should therefore reason that

Hypothesis 2. *A focal firm's 2-Ego-Network Centrality, ceteris paribus, is larger than or equal to it's 1-Ego-Network Centrality.*

Recall that *clustering* is the construct that indicates "how close" or "how tight" a group of firms are to each other. Clustering is essentially a characteristic of a group of firms "working close together and with each other", or, put another way, the extent of forming a social "clique". In the 2-ego-network level, in order for the property of cliques to form, there needs to exist, at the minimum, a high number of triads in the 2-ego-network. While it can be expected that a higher level of triads exist in 2-Ego-Networks, the number necessary to create large cliques also increases with the number of nodes that a firm in the 2-Ego-Network is connected to.

In addition, there is only one instance where clustering would be high in a 2-ego-network: the firms not in the ego-network that are in the 2-ego-network would all need to be connected. Since a firm can have well over 200 suppliers, it would be unreasonable to observe this level of forming cliques with this many of nodes. That is, greater efforts from a greater number of firms would be required to form a clique. This does not seem reasonable. Those efforts are not as high at the ego-network level, however, since there will most likely be fewer nodes in the ego-network. Put another way, it is easier for a firm within a 1-ego-network to form a clique than it is within it's 2-ego-network. Therefore, we should expect to see

Hypothesis 3. *A focal firm's 2-Ego-Network Clustering, ceteris paribus, is smaller than or equal to it's 1-Ego-Network Clustering.*

Since the 2-Ego-Network is at least as great in size as the 1-Ego-Network, the 2-Ego-Network will most likely have a higher probability of containing a circular path (though not necessarily triads). However, the cycles will be of a greater size than those

of the 1-Ego-Network (since there are a greater number of nodes). Therefore, we should expect

Hypothesis 4. *A focal firm's 2-Ego-Network Feedback/Thickness, ceteris paribus, is greater than or equal to it's 1-Ego-Network Feedback/Thickness.*

Last, again, since there are a greater number of nodes in the 2-Ego-Network, the number of arcs in the 2-ego-network should be expected to be at least as great as those in the 1-Ego-Network. The effect of removing a node or an edge in the 2-Ego-Network should not be as damaging as removing one in the 1-Ego-Network if there are a large number of arcs. However, the primary issue with the 2-Ego-Network is that it's size will generally be much greater than it's ego-network counterpart. We would need to observe a proportional change in the number of arcs within the 2-Ego-Network. While the 2-Ego-Network will in fact contain more nodes than those of the Ego-Network, the number of arcs do not grow as fast, since many are left out into the 3-Ego-Network. That is to say, node density should be expected to be quite low. Given the higher potential for firms to cluster within the ego-network rather than within the 2-ego-network, it would be reasonable to presume that the 2-Ego-Network is more sparse. That is to say,

Hypothesis 5. *A focal firm's 2-Ego-Network Connectivity, ceteris paribus, is less than or equal to it's 1-Ego-Network Connectivity.*

Another inquiry that comes to mind is just how might the structural dimensions of a 2-Ego-Network affect the 2-Ego-Network's risk? There is some precedent to base my arguments on. Choi and Krause had argued that the "complexity" of a supply base is in a positive quadratic association with a focal firm's supply risk (Choi & Krause, 2006). However, "complexity", as they defined and used the construct, is a cumulative construct of the network structure. He does however further break this down into three sub-constructs, namely the number of firms, the level of differentiation, and level of inter-relationships. He argues that the level of differentiation is in a positive curvilinear association while the level of inter-relationships is in a negative curvilinear association with risk when the number of firms is held constant (Choi & Krause, 2006).

They define *differentiation* in regards to contextual characteristics of the firm, but implies that such levels of differentiation are essentially manifested into a structure, such that the "structure refers to the group of companies that have organized themselves around a powerful focal company and work together as if they belong in the same clan" (Choi & Krause, 2006). On the other hand, they define *inter-relationships* as the tendency for firms to collaborate and form triads. He argues that firms of whom are less differentiated expose themselves to less risk (while firms with less inter-relationships have higher levels of risk, due to lack of ability to handle disruptions and a loss in opportunity to learn and share information).

Essentially, Choi and Krause has argued, albeit on a higher dimensional level, that risk is in a non-linear relationship with structure. This view has been supported by others in the literature as well. Kim argues that as networks exhibit power-law degree distributions (few nodes with high levels of degree, many nodes with low levels of degree), the risk (they use resilience, but their definition resembled robustness, a sub-construct of risk), tends to be lower (although, not necessarily linear) (Y. Kim et al., 2015). Mildly contradictory to Choi, Bode had found that there was a positive-curvilinear (not necessarily quadratic) relationship between the supply chain complexity and the risk (Bode & Wagner, 2015). Basole had found similar conclusions with his own "health measures" in networks that exhibited small-world properties (Basole & Rouse, 2008).

Therefore, the literature seems to have lightly and within an ad-hoc manner, established that there exists a non-linear relationship with the construct of supply chain risk (it should be emphasized here again then when I refer to "risk" in this dissertation, I am referring to "risk propagation", something the extant literature has not studied in depth). Where the extant literature lacks, however, is in (1) describing the relationship between more refined constructs of network structure and risk and (2) describing such a relationship with risk propagation (which is a distinct construct from just risk).

Choi's work is the closest that comes to breaking down complexity of the structure of a network into smaller pieces, but relies on some contextual factors in doing so, and merely treats sub-dimensional structure as an *a posteri* effect. Choi's conceptualization

of *differentiation* is most closely related with the structural dimensions of Distance and Connectivity, while his conceptualization of inter-relationships are more aligned with the dimensions of Clustering, Centrality, and Feedback. Choi's constructs of differentiation and inter-relationships can be mapped to the 5 Fundamental Dimensions that I have described above. As such, I will be able to formulate my hypotheses by leveraging a combination of theory designed by Choi, Bode, and Basole from the extant literature.

Choi's definition of his differentiation dimension is primarily based on the idea that firms can either decide to cluster together with other agents of whom share similar philosophy and culture, or, they can choose to differentiate themselves and instead work with many individuals, each of opposing cultures and philosophies. Given two networks of the same size (same number of nodes), if one network has a larger distance than the another network, then this is indicative of the former network being more "spread apart". That is, on the whole, most nodes are not tightly formed together, but rather are connected within a very sparse manner. As such, networks that exhibit this type of distance-based sparsity could be argued to be a reflection of a network that is "differentiated". It would be difficult to argue otherwise, for if a network were "sparse" or "spread apart", generally speaking, then by definition, there exist fewer connections within the network (more connections would imply that the shortest distance between two points could decrease). This leads to the conclusion that the dimension of Distance would be positively associated with the construct of "differentiation". So we should therefore anticipate that

Hypothesis 6. *A focal firm's 2-Ego-Network Distance, ceteris paribus, is in a positive curvi-linear relationship with 2-Ego-Network Risk.*

The construct of connectivity describes the network's "robustness". That is, it is a characterization of the network's ability to remain connected in the event of some form of removal (node, edge, group of nodes/edges). Highly connected graphs would imply that the number of unique pathways between any two nodes is generally higher than a network of the same size who's connectivity is less. That is, less connected graphs require the removal of fewer members in the network in order to disconnect the graph.

This in itself would imply that fewer distinct paths run between the various nodes in the network. Given the level of paths between any two nodes in the network, this would transcend the notion that nodes are more "spread out", not in the sense of distance, but rather that of ability to travel.

Logic would dictate that highly clustered groups of individuals would not be as well connected as less clustered networks. If a network has many tight clusters of groups (that is, high levels of clusters), then few actors will only interact with a few other actors. It would indeed be rare, and run counter to definition, for an actor that is not well differentiated to have many connections to actors within different cliques. In other words, taking the "clique" as the unit of analysis itself, we should expect to see high levels of connections within the clique, but, very low levels of connections between the cliques. As such, the connectivity construct would be positively associated with the differentiation construct (higher levels of connectivity should exhibit higher levels of connectivity). Hence, we can speculate that

Hypothesis 7. *A focal firm's 2-Ego-Network Connectedness, ceteris paribus, is in a positive curvi-linear relationship with 2-Ego-Network Risk.*

Choi's conceptualization of inter-relationships is primarily concerned with describing a firm's willingness to move from a dyadic perspective in buyer-supplier relationship management towards a triadic viewpoint, as well as other firms within a potential triad strengthening it's relationships with each other. Networks that display higher levels of inter-relationships, between and within firms, should be expected to exhibit many "well-connected" firms within that network that are themselves "well connected". Hence, for a network where centrality levels tend to be fairly high, then we should be able to expect the inter-relationships of the to also be high. As a firm becomes more central, it's accessibility to other nodes within the network because elevated, which would translate to a decrease in it's distance to other node. Hence, it would be natural that it follows that

Hypothesis 8. *A focal firm's 2-Ego-Network Centrality, ceteris paribus, is in a negative curvi-linear relationship with 2-Ego-Network Risk.*

Networks where clustering is fairly high, by the previous argumentation, should be expected to have smaller distances as well as smaller levels of connectedness. As such, it would naturally follow that network with high clustering would be associated with networks where inter-relationships are high. When cliques form within a network, a firm within the clique would be more highly accessible, and can be expected to interact more with the other actors within the clique (by definition, if it's a click, then there must be a high level of interaction of some nature between the firms within the clique). Hence, it would naturally follow that

Hypothesis 9. *A focal firm's 2-Ego-Network Clustering, ceteris paribus, is in a negative curvi-linear relationship with 2-Ego-Network Risk.*

Last, when the feedback of a network is high, then it is illustrative of the fact that there are either many cycles, cycles of larger size, or both. In the case of the former, a network that contains many cycles within the network, when compared to a network of the same size with fewer cycles, should be expected to have more triads and higher clustering coefficients. In the latter case, large cycle sizes would imply a high degree of clustering, since many triads would be necessary in order to form a cycle of large size in the first place. Hence it naturally follows that networks of which possess higher feedback levels would also be expected to exhibit higher inter-relatedness levels. Therefore, it should follow that

Hypothesis 10. *A focal firm's 2-Ego-Network Feedback, ceteris paribus, is in a negative curvi-linear relationship with 2-Ego-Network Risk.*

Further, it has been established in the literature that structural characteristics at different tiers in the supply network affect risk differently (Bode & Wagner, 2015). The notion of the "tier" of the supply network is a conceptualization of Lambert's notion of horizontal distance (Lambert et al., 1998). It can be argued that the structure of a k-ego-network is itself a generalization of Lambert's notion of a "tier". His construct is only encompassing of a single direction (either downstream, or upstream), that is, the upper tiers or the lower tiers. The k-ego-network construct is encompassing of both directions and outward from a fixed position. For example, the 1-ego-network

would include the 1-tier suppliers and the 1-tier customers. The 2-ego-network would encompass the 2-tier suppliers and the 2-tier suppliers (as well as the 1's). That is, the k-ego-network is a cumulative structure of tiers.

1-Ego Network Risk

I have hypothesized that the overall structure of a 2-Ego-Network should be expected to be different than that of the 1-Ego-Network. As such, this is partial motivation as to how the structure of a 2-Ego-Network might portend different risks on such a network than those of the lower 1-Ego-Network. Indeed, it has been found in the past that firm's exposed to high risk nodes are themselves at risk (Basole & Bellamy, 2014b). Increasing the network view from ego-network to 2-ego-network would imply that a firm of high risk in the 2-ego-network that is not in the 1-ego-network could potentially transfer that risk to the focal firm via the ego-network structures. That is, there is a greater possibility of high-risk firms in the 2-ego-network than there are in the 1-ego-network.

It can further be argued that firms in the 1-ego-network, depending on the type of collaboration strategies employed, are more generally aware of each other's risks via information sharing. As such, risk should generally be lower in the 1-Ego-Network as a result. However, a focal firm will not have interaction with a firm in it's 2-Ego-Network. A decrease in visibility and information could lead to a loss in identifying proper mitigation strategies. Hence, since the structure of the Ego-Network and 2-Ego-Networks are different, and, since the risk distributions of the networks would most likely differ as well, we should anticipate that:

Hypothesis 11. *A focal firm's 2-Ego-Network Risk, ceteris paribus, is greater than or equal to it's 1-Ego-Network Risk.*

As I mentioned earlier, there is some preliminary evidence that structure of supply networks affect risk at different tiers in the network (Bode & Wagner, 2015). Since the k-Ego-Network structure is a generalization of the more traditional "tier" structure, it would be logical to infer that structure affects risk different at varying degrees of

k-ego-networks. Choi's original conceptualizations of network complex and their associated relation with risk was primarily discussed at the network level. As such, I argue that Choi's conceptualization of differentiation and inter-relatedness is analogous to the 2-Ego Network. The 2-Ego-Network of a firm would be sufficiently large enough to consider application of Choi's constructs. However, I argue that Choi's conceptualization of "network structure" and "risk" is not applicable in the same manner to Ego-Networks or Local-Risk.

When a firm is engaged with other firms in their ego-network (by definition of course), that network will of course exhibit certain structural properties, as I have argued previously. These structural properties are inherently dependent on the level of ego-network in which they are considered. As I argued earlier, the size of a 2-ego-network will be larger than the size of a 1-ego-network (with the exception of a few trivial cases). Choi's theory on the relationship between complexity and risk could be argued to be more reflected within networks of larger size. Indeed, as far as this author is aware, these properties are reflected in higher-tier levels of Bode's simulation study (Bode & Wagner, 2015).

When a node is added into a firm's ego-network, the size of the ego-network changes by a factor of less than 1. However the effect on the size of the firm's 2-ego-network can be much larger than this. If a focal firm has 200 firms in it's ego network, and 400 firms in it's 2-Ego-Network, and decides to add a firm of whom has in their own ego network 500 distinct firms that are not in any way associated with firms in the focal firm's ego-network, then the size of the ego network increased by a factor of 1.005, while the size of the 2-Ego-Network increased by a factor of 2.25, a percentage difference of 124%. In other words, for every 1% change in the Ego-Network, there is a 250% change in size of the 2-Ego-Network (assuming uniqueness of firms).

It would be logical to presume that when an Ego-Network changes by a small amount, the impact to the ego-network risk is minimal, as not much about the overall network structure has changed. However, when a 2-Ego-Network changes, it does so not in small quantities, but in large quantities. That is, 2-Ego-Networks change as a result of changes in 1-Ego-Networks. Such changes to the 2-Ego-Networks would be

large, as I demonstrated earlier. Therefore, it would be logical to assume that typical changes to the 2-Ego-Network would result in large changes in risk as a result, either through the introduction of new nodes, or new edges that were not previous considered to be part of the 2-Ego-Network (but now are due to the changes). Therefore, just due to as a matter of size differences, we should anticipate at least some differing effects on 2-Ego-Network risk from the structure.

When the 2-Ego-Network changes, as I have argued, it will most likely change in large ways. The impact of 1-Ego-Network structural changes, however, should be anticipated to be minimal, since the changes to the 1-Ego-Network structure is minimal. Furthermore, it can be argued that probabilities of disruption within Ego-Networks should be anticipated to be generally lower than those in 2-Ego-Networks. This is due to higher-levels of information sharing and collaboration and coordination efforts of those within the focal firm's ego-network. Observing this strictly from a structural perspective, when the sizes of the Ego-Networks are small (as most are when compared to the sizes of 2-Ego-Networks), the levels of connectivity and distance will be lower, while the levels of clustering, feedback, and centrality would be larger, due to the small size of the network. As such, in 1-Ego-Networks, I do anticipate similar affects to risk, however, I will note that I hypothesize these affects will will be on a smaller level than the effects within the 2-Ego-Network. Hence, we should expect:

Hypothesis 12. *A focal firm's 1-Ego-Network Centrality, ceteris paribus, is in a negative curvi-linear relationship with 1-Ego-Network Risk.*

Hypothesis 13. *A focal firm's 1-Ego-Network Clustering, ceteris paribus, is in a negative curvi-linear relationship with 1-Ego-Network Risk.*

Hypothesis 14. *A focal firm's 1-Ego-Network Feedback, ceteris paribus, is in a negative curvi-linear relationship with 1-Ego-Network Risk.*

Hypothesis 15. *A focal firm's 1-Ego-Network Distance, ceteris paribus, is in a positive curvi-linear relationship with 1-Ego-Network Risk.*

Hypothesis 16. *A focal firm's 1-Ego-Network Connectedness, ceteris paribus, is in a positive curvi-linear relationship with 1-Ego-Network Risk.*

Node/Arc Level Risk

When we observe the network from the local perspective, that is, from the individual firm or relationship level of analysis, we are able to characterize the structure of that individual firm/relationship based on it's structural properties within the overall network itself. In some sense this is a paradox. We are observing "local" properties, but those local properties are inherently defined based on the "global" structure of the network. None-the-less, these individuals in the network exhibit these structural properties, and one can argue that "how they lie" within the overall network, from the 5-Dimensional perspective, will portend it's inherent risk.

For example, when the centrality of a node (with respect to the entire network) is high, then it is highly accessible to other nodes within the network. That is, there are many paths of relative short distance between the location of interest (node or edge) to another location (again, node or edge). It would be logical to presume that when the node is highly accessible to a variety of other nodes within the overall network, that it puts itself at risk due to the risk of those other nodes. Unlike 2-Ego Networks and Ego-Networks that will have a smaller number of nodes (and hence, lower levels of centrality), node/edge based centrality will generally be higher.

Choi has argued in that with respect to supply bases (partial ego-networks), that higher centrality will portend higher levels of risk. We should anticipate that since node/edge-level characteristics are always considered with respect to the global network, then their various dimensional characteristics should be "larger" than those of their k-Ego-Network counter-parts (since "Ego-Network" Dimensions would only consider the structural attributes within that induced network, and not the entire network).

This may seem paradoxical due to the natural ordering of the levels of analysis (2-Ego-Network > 1-Ego-Network > Local). However, the ordering, based on the definition of the structural characteristics with respect to the level of analysis, would actually be different (Local > 2-Ego-Network > 1-Ego-Network), since the local characteristics are always considered from the global network perspective, which is distinct from how the k-Ego-Network characteristics are considered. Hence, if we argue that the structural

constructs at higher levels of analysis follow the reasoning of Choi, of which I do here, then we should see the same patterns, yet possibly of different degrees, within the relationship between local structure and risk. Therefore, we should anticipate by the reasoning I have presented above that

Hypothesis 17. *A focal firm's Local Centrality, ceteris paribus, is in a negative curvi-linear relationship with Local Risk.*

Hypothesis 18. *A focal firm's Local Clustering, ceteris paribus, is in a negative curvi-linear relationship with Local Risk.*

Hypothesis 19. *A focal firm's Local Feedback, ceteris paribus, is in a negative curvi-linear relationship with Local Risk.*

Hypothesis 20. *A focal firm's Local Distance, ceteris paribus, is in a positive curvi-linear relationship with Local Risk.*

Hypothesis 21. *A focal firm's Local Connectedness, ceteris paribus, is in a positive curvi-linear relationship with Local Risk.*

A point of contention here is that I have mentioned "local risk" without regard to the definition of"local", that is, "edge" or "node". I contend that both node-base and edge-based measures affect the local risk in the same manner. However, I will hypothesize one more point. Nodes and edges are distinct in that a node is self-contained, while an edge always requires the definition of two nodes to exist (it makes no logical sense to speak of an edge that only has "one" connection to a node). However, what may be of interest is how the "edges" are structurally related to other "edges", and, if this structure portends effects on node/arc risk. Due to the fact that supply networks will ultimately have more connections than actors, it would be logical to presume that the affects of "edge structure" would impact risk at a more heightened effect than "node structure" would. This is just simply due to the number of edges in the graph. I speculate the form and direction of the relationships between edge and node structure and edge/node risk are the same as I have hypothesized above (as "local"). I will make one clear hypothesis however, given this logic of "more edges" than "nodes":

Hypothesis 22. *A focal firm's edge structure has a greater impact on edge risk than the focal firm's node structure has on node risk.*

Cross-Level Risk Effects

A fairly understudied topic in the literature is how risk "propagates" from one (structural, not supply chain) level of analysis to another. Only a handful of studies have considered studying risk at the tier level (the ones I have mentioned previously). However, to the best of this author's knowledge, it has yet to be understood how risk "flows" from higher-order groups of nodes to lower-order groups of nodes (ie from k-ego-network to (k-1)-ego-network), if it flows at all. An analogy is in order here to understand this embedded and cumulative effect of so called cross-level risk flow.

During the financial crisis of 2008, the United States had suffered one of the worst recessions in nearly a generation. It has been speculated that one of the primary sources of the "disruption" to the economy was primarily due to the banking industry, as a result of underwriting loans to individuals of whom could not inherently afford the credit they were taking. This of course was due to government regulation via the Community Reinvestment Act, which guaranteed financial institutions their backing of the loans in the event of default. The lack of proper liquidity requirements to maintain a stable financial system had essentially led to a full collapse of the private-equity markets. As a result, many found it difficult to gain proper credit to finance their various economic activities. This eventually "trickled down" to various industries, of whom could not acquire the proper levels of operating cash to finance their day-to-day activities. The result was massive layoffs across the economy. As such, at the individual level, the typical working household had to rely on unemployment and government assistance as well as savings until they can later find meaningful employment.

The previous historical anecdote illustrates how risk emanates at a higher level of analysis (the general American economy) down to the smallest level of analysis (the household, or, individual actor within the greater economy). Similar arguments are often posed for/against massive tax cuts at the Federal level in terms of these actions "propagating" throughout various levels of analysis. Risk behaves no different in supply

networks. We could arguably notice that a sub-network (such as a k-ego-network) has been "disrupted". From this level, it can be argued that the disruption "flows" down to a smaller level of analysis, say, the firm level, and of course subsequently down to the smallest level of analysis (the "people-supply chain", which is not considered in the current arguments).

The primary question at hand is, how does this notion of "risk flow" manage to move from a very global and abstract environment down to a very local and deterministic environment? I argue that one way that this flow can be characterized is through the leveraging of the k-ego-construct. Disruptions need not occur at the "network-level" in order to eventually "tickle down" to the local level. They can occur at any level. The key importance to the analysis here is of course first fixing a point of reference, just as we do in geometry. If we fix the point of reference to be the focal firm, then disruption can logically occur at any location in the network. It can be argued that viewing sources of risk from this perspective is only viewing risk on a projected dimension of a much larger space of analysis.

Rather, in reality, "risk" manifests relative to a fixed focal firm at a specific "k-ego-network". The risk then "spreads down" from the k-ego-network to the (k-1)-ego-network, and so forth, until it spreads, or stops, at a an $n < k$, or, at the focal firm, of whom will either be able to thwart the disruption, or, either intentionally or unintentionally enable it to continue it's path of destruction, ironically, back downwards to the outer k-ego-networks. Therefore, risk "spreads" in regards to (1) a point of reference (the focal firm) and (2) through various nodes that are members of a variety of embedded dimensional networks (i.e. the k-ego-networks). We can think of this as somewhat of a "Theory of Relativity for Supply Network Risk Propagation":

Hypothesis 23. The Theorem of Relativity for Supply Chain Risk Propagation
Suppose a focal firm has at most n-ego-networks. Then if a disruption is to manifest at the $k \leq n$ level of analysis, then the risk of disruption, but not necessarily the disruption itself, will spread to the $(k-1)$-ego-network level.

This is a very general theorem, of which will require careful additional examination

in a future study. However, the general spirit of the theorem will be studied in this dissertation by observing the effects of 2-Ego-Network Risk on 1-Ego-Network Risk. I have argued that risk can spread, relative to a focal firm, from higher-order ego-networks to lower-order ego-networks. The natural extension to this line of theory is to inquire as to how the risk may spread from one level down to the next? In this dissertation, I will explore the linear effects of spread of risk between levels. It would be worth while in a follow up study to determine more carefully how risk spreads, and if it spreads in more than just a linear manner.

Hypothesis 24. *The 2-Ego-Network Risk, ceteris paribus, is in a positive linear relationship with the 1-Ego-Network Risk.*

Logically, by induction, we should be able to infer that the 1-Ego-Netowrk risk is in the same relationship with the local risk. Hence, I suggest that

Hypothesis 25. *The 1-Ego-Network Risk, ceteris paribus, is in a positive linear relationship with the Local Risk.*

Furthermore, I would like to argue that, again, the risk at the 1-ego-network level would most likely affect the node risk differently from the arc risk. This may be due to the fact that an edge requires two nodes, while a node does not require any such connections. The risk that affects the edge is a reflection of the agreement between two parties and their perspective on the edge's own ego network (the union of the ego-networks of the two firms). Therefore the ego-network risk of an edge is naturally larger than that of a node's. I will argue, however, that regardless of node or edge, the same relationship holds. Namely, that it is a positive quadratic relationship. However, I argue that the edge's impact of the edge's ego-network risk on it's own risk is more substantial than that of the node. That is, there is more opportunity for the edge to engage in "risk compensation", since the compensation is a collective matter in this case. If both sides of the edge are fairly low-risk-averse, this will then attenuate the curve much higher than the node's ego-network risk on the node-based risk.

In any case, I have demonstrated here that the structure within a structural level of analysis impacts the risk within that level of analysis. In addition, I have demonstrated

that these relationships are anything but homogeneous, but rather, are inherently dependent on the particular level of analysis that is being studied. In addition, I have demonstrated that the structural differences across the different levels of analysis is in part to explain the differences in risk across these levels.

Chapter 4

Research Design

4.1 Population and Variable Definition

4.1.1 Description of the Population of Interest

The population of interest in this research is the collection of all supply networks, their respective embedded 2-Ego-Networks and 1-Ego-Networks, as well as the collection of all nodes and edges embedded in those networks, with respect to no specific level of Supply Chain Level of Analysis. The goal of this research is to understand how the structure of the various "sub-networks" impact the "sub-network's" risk. As such, this dissertation focuses on the population of all supply networks.

Supply networks, as I have previously covered in the literature review, tend to follow a "hierarchical" system, although not always. It recently has been suggested that supply networks actually follow more of a complex adaptive system (Carter et al., 2015). In either case, it can be argued that the network indeed follows at least some form of hierarchy (although, very loosely) in terms of "raw materials", "manufacturers", "distributors", and "retailers". The primary assumption regarding the structure of the networks considered is that they are "evolved", starting a discrete number of "focal firms", and building upwards and downwards to other firms, at random. While the resulting network may be "random" or "complex", the heart of the structure can still be characterized by the number of "levels", although I opt not to do so here, since the primary inquiry of this dissertation is to focus on core-graph-theoretic constructs and measures.

4.1.2 Dependent Variables

The primary dependent variables in this research were 2-Ego Network Risk, 1-Ego-Network Risk and Node/Arc Risk. Risk was characterized using the three sub-constructs mentioned earlier (contribution, significance, and velocity), and they were all operationalized in three ways (average, standard deviation, and coefficient of variation). First, from (Garvey et al., 2015) model using the Expected Location Risk Contribution Factor (ELRCF) (discussed earlier). The second conceptualization leveraged a new measure I designed, of which I call the Expected Location Risk Significance Factor (ELRSF). This is a diagnostic measure that seeks to measure the number of anticipated ancestors that we can expect to have been disrupted given that a risk at a location had manifested or not manifested (knowledge of disruption/no disruption is known). It's definition is:

$$ELRSF(n) = (1 + ERSF(n|n = 1))P(n = 1) + ESSF(n|n = 0)P(n = 0)$$

where ERSF is the Expected Risk Significance Factor, which is defined for a node n assigned to a value $a \in \{0, 1\}$ and n has a set of ancestors $A(n)$:

$$ERSF(n = a) = E\left[\sum_{i \in A(n)} U(n_i)|n = a\right]$$

The last construct of the dependent variable is yet measured a new way in which I have termed the *Expected Location Risk Velocity (ELRV)*. This measures the average percentage change in the risks present downstream compared to risks upstream. Negative values would indicate that the current node is a "depressor" of risk, while positive values would indicate an "increaser" of risk (analogous to increasing "position" or "decreasing" position change). The mathematical definition would be:

$$ELRV(n) = ERV(n|n = 1)P(n = 1) + ERV(n|n = 0)P(n = 0)$$

where ERV is the Expected Risk Velocity, which is defined for a node n assigned to

a value $a \in \{0, 1\}$

$$ERV(n = a) = E\left[\frac{\sum_{i \in D(n)} U(n_i) - \sum_{i \in A(n)} U(n_i)}{\sum_{i \in A(n)} U(n_i)} | n = a\right]$$

Given the different constructs of risk, one must beg the question "which shall we use to test the hypotheses?". These are all sub-constructs of the primary risk construct. Recall our discussion earlier. Risk significance is the construct of "pre-risk prorogation" (what has happened already?). Risk confidence is what is anticipated to happen further down the stream. I argue that Risk Velocity is the most appropriate and most related to the general risk construct. The velocity measures the percentage change of risk flow. Hence, negative values indicate that the node is a "risk depressor", while positive values indicate that the node is a "risk enabler". Hence, from the perspective of the node (or more generally the edge), being "risky" in the sense of increasing one set of damages into another (or depressing them), the risk velocity would be the most appropriate sub-construct. The study still controls for the significance and the confidence out of both curiosity and proper statistical control of latent factors. But the primary variable used for the hypothesis testing will indeed be the velocity. It's specific operationalization, the coefficient of variation, will be used as the primary dependent variable to use to formally conduct the hypothesis testing.

4.1.3 Independent Variables

The initial independent variables considered in this research were mainly the structural variables of a network. I will later discuss how I have reduced these down to the five conceptual dimensions that I had discussed earlier. I considered the following independent variables in the current research, all of which were calculated using R and the *igraph* package. The First and Second Ego-Network levels share the same structural variables. For the first and second ego-network level, the independent variables considered were:

1. Number of Nodes/Arcs - The number of nodes in the 2-ego and 1-ego networks, respectively.

2. Edge/Node Connectivity - The number of edges/nodes that must be removed from the 2-ego or 1-ego network, respectively.

3. Diameter/Radius - Diameter is the maximum of the eccentricities of the nodes (the maximum geodesic distance from the current node to any other node). The radius is the minimum of such values. Namely:

$$diameter = \max_{i \in N}\{\epsilon(i)\}$$
$$radius = \min_{i \in N}\{\epsilon(i)\}$$

4. Girth - The smallest length cycle in the graph. It is set to 0 if there are no cycles in the graph.

5. Clique Number - A *clique* is a collection of nodes in a graph such that the induced subgraph forms a complete graph. The Clique Number is the largest number of nodes that form a clique.

6. Number of Largest Cliques - This is a count of the number of cliques that have the size of the clique number of the graph.

7. Average Clique Size - This is the average of the sizes of all unique cliques in the network.

8. Number of Structural Holes, Number of Triads - A structural hole is a set of three nodes such that two nodes share a common neighbor, but are not neighbors themselves. A triad is a complete graph of size three. These measure the total number of unique structural holes and the number of unique triads in the network.

9. Centralized/Average Constraint - Constraint is a measure of "cluster". It essentially reflects the level of neighbors that "influence" the current node in such a way that "prevents" it from connecting with new nodes. It is a common measure of "influence" in social network theory. I omit the formula here for complexity and scope issues. The constraint of a network can be calculated in two ways: Freeman's Centralization (to determine how "central" the node is with respect to constraint) and a typical mean.

10. Density - The density of a network is the number of arcs divided by the the number of possible arcs the network can have. The density of a node is the proportion of it's degree to the total number of possible nodes it can connect to in the network.

11. Centralized/Average Clustering Coefficient - The clustering coefficient is the proportion of actual triads a node exhibits with it's neighbors to the total number of possible triads a node can exhibit with it's neighbors. This is a local-based measure. It can be generalized to a network-based metric by either (1) calculating the Freeman Centrality of the Clustering Coefficient or (2) finding the mean of the Clustering Coefficient.

12. Average Shortest Path Distance - This is the average of the shortest paths between two any two nodes. Namely, if $d(i, j)$ represents the geodesic distance of the shortest path between node i and node j, then:

$$ASPD = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} d(i, j)$$

13. Centralized/Average Betweeness Centrality - This measures the proportion of paths from any two nodes that pass through a given node. Namely, given a node k:

$$BC = \sum_{i \neq k} \sum_{j \neq i, j \neq k} \frac{\sigma_{i,j}(k)}{\sigma i, j}$$

where $\sigma_{i,j}(k)$ represents the number of unique paths between i and j that pass through k, and $\sigma_{i,j}$ represents the number of unique paths between i and j. This measure can be measured at a network level by finding either the Freeman Centrality or the mean.

14. Centralized/Average Eigen Centrality - This measures how "well-connected" a node is to other "well-connected" nodes. It can be computed by looking at the eigen-centralities of it's neighbors. If we would like to compute the eigen-centrality for node n, and $N(n)$ is it's neighbors, then we can compute n's eigen-centrality

by:

$$e_n = \frac{1}{\lambda} \sum_{k \in N(n)} e_k$$

The measures can be found by computing the eigen-vectors of the adjacency matrix of the entire network (assuming it is undirected). The vector of centralities would be the absolute value of the eigen-vector that corresponds to the largest eigen-value.

15. Centralized/Average Closeness Centrality - This is a measure of how "close" two nodes are from each other. It is the opposite of the "average shortest path". That is, the closeness centrality of a node n, with $d(n, i)$ representing the shortest path between nodes n and i, is:

$$C(n) = \frac{1}{\sum_{i \neq n} d(n, i)}$$

Once again, this can be generalized to a network-measure by finding the Freeman Centrality or the average of the values of all nodes.

16. Centralized/Average Degree Centrality - The degree centrality of a node is the size of it's collection of neighbors. That is:

$$DC(n) = |N(n)|$$

This can be generalized to a network-measure by finding the Freeman Centrality or the average of the values of all nodes.

17. Centralized/Average Eccentricity - For a node n, it's eccentricity is the largest geodesic distance (the shortest path to a node i, $d(n, i)$) to any other node i in the network. Namely:

$$\epsilon(n) = \max_{i \in N, i \neq n} d(n, i)$$

This can be generalized to a network-measure by finding the Freeman Centrality or the average of the values of all nodes.

18. Graph Center - While this is not a graph measure, per se, it is used for other measures. The graph center is the collection of nodes with the minimum eccentricity. That is:

$$\omega = \{n \mid \epsilon(n) = \min_{i \in N}\{\epsilon(i)\}\}$$

19. Min Distance to the Center - The minimum distance of a node n to the center of the graph is smallest distance to a node that is in the center. Namely:

$$MinCenter(n) = \min_{k \in \omega}\{d(n,k)\}$$

20. Max Distance to the Center - The maximum distance of a node n to the center of the graph is the largest distance to a node that is in the center. Namely:

$$MaxCenter(n) = \max_{k \in \omega}\{d(n,k)\}$$

21. Coreness - The coreness of a node n is the level k such that the node n is in the k-core of the graph, but not in the $(k+1)$-core of the graph. The k-core of a graph is the induced subgraph of nodes such that all the nodes in the k-core have a degree of at least k.

22. k-Ego-Connectivity - The smallest value of k such that the k-ego-network disconnects the graph.

All of the variables described here either had a canned method in the *igraph* package of R, or, were hand-coded by the author of this dissertation. There are three levels of analysis in this dissertation, namely the 2-Ego-Network level, the 1-Ego-Network level, and the local level. The measures were calculated for each level using the method described earlier. The 2-Ego-Network and 1-Ego-Network are levels of analysis that involve multiple nodes and edges. Hence, the metrics were computed using either (1)

network-based metrics (2) the Freeman Centralization and (3) the average. Node-level attributes were computed directly according to their respective definitions. The edge-level attributes were computed by first finding the *line graph* of the graph, and then computing node-level metrics. The *line graph* of a graph is the resulting graph by assigning nodes to represent the edges in the original graph, and by connecting two nodes if their corresponding edges in the original graph were incident (that is, they share the same vertex in the original graph).

Control Variables

The study controlled for three primary variables of interest, namely the *alpha* parameter, the *beta* parameter, and a dummy variable indicating if a data point in the local data was a node or an edge. The alpha and beta needed to be controlled for since probability obviously is a big factor in the calculation and determination of the risks. However, controlling for all probabilities would be cumbersome. Hence, one way to control for the probabilities simulated is to simply use the parameters from the distribution in which the probabilities were sampled, as a regressor in the primary model, so that proper structural effects can be isolate and omitted variable bias can reduce.

4.2 Design of the Simulation

The simulation was designed to sample different supply network structures of varying sizes and structure in order to capture the variability in structure so that if a relationship between structure and risk exist, the relationship can be properly isolated. The simulation operated off of the following assumptions:

1. A network had a minimum of 100 nodes.
2. The network had a variety of "levels", as small as 5.
3. A node in the network had at most 10 parents.
4. Every location in the generated network corresponded to one and only one "risk". This is a bit of a departure from Garvey's model. However, it simplified the

calculations. Hence, every node and edge that exists in the generated supply network had one and only one risk associated with it.

5. The resulting *risk graph* was generated according to (Garvey et al., 2015) algorithm. Only parents that were generated for a node were allowed to connect to their direct children, or edges connecting to the children, and no others.

6. Edges in the supply network were assumed to connect to two node locations from the supply network. The risk at the edge was represented as a node in the risk graph. It was only allowed to have one parent, and one child.

7. The conditional probabilities of the distributions in the risk graph were sampled randomly from the same Beta Distribution with a given α and β value. That is, an α and β value chosen at random (uniformly), and all the conditional probabilities in a given risk graph were sampled from the same $Beta(\alpha, \beta)$.

The design of the simulation involved generating the supply networks and their corresponding risk graphs. From there, their respective node/edge-level risk measures were computed. After, for each node and edge in the network, the 2-Ego-Network was found, and structural characteristics of only the 2-Ego-Network were recorded, along with the risk metrics for the 2-Ego-Network (an average, standard deviation, and coefficient of variation of the respective risk measures of the nodes and edges in only the 2-Ego-Network). The same was conducted for the 1-Ego-Network. The node-level attributes were computed using the same package, and the corresponding risk measures for the node were recorded. For the edges, the line graph was used as the primary mechanism to calculate the edge's structural characteristics. The risk measures for the edges were also recorded.

The simulation was conducted using a custom-engineered system of virtual machines leveraging a variety of Amazon EC2 Instances, with an R-Instance installed on each of them. 10 smaller machines were used to generate the networks and compute their risk graphs. Every 10 networks that were generated were stored into a single compressed file using the Brotli compression algorithm. The file was then stored on an Amazon

S3 Bucket. A second phase of the simulation (which ran in parallel to the first phase) was to have three higher-quality Amazon EC2 instances "listening" on each of the S3 Buckets.

As soon as a file was placed on there from the first phase, the EC2 instance would sweep it up, decompress it (again, using Brotli), and compute the risk measures specified above using *bnlearn*' s likelihood-weighting simulation belief inference algorithm. After this machine was complete with conducting the full computation of the risk scores, it saved the networks into a file and again compressed it using the Brotli algorithm. The resulting file was then placed automatically on a separate S3 Bucket. A third tier of the simulation had a separate three Amazon EC2 instances listening on these 2nd tier buckets. As soon as a file was placed on these buckets, the code would sweep up the file, decompress it, and calculate (1) all the possible 2-Ego-Networks in the graph, (2) all the structure metrics for each 2-Ego-Network and risk scores, (3) all the possible Ego-Networks in the graph, (4) all the structure metrics for each Ego-Network, (5) all the node and edge metrics along with their corresponding and respective risk scores.

The final results were placed into three different tables (one for the 2-Ego-Network data, one for the 1-Ego-Network Data, one for the edge/node data). The file was then again compressed using Brotli and moved to a different S3 Bucket. The last phase, a final EC2 instance was listening on the last tier of S3 buckets for new files. The instance would read the file, decompress it, and append a data frame. Once the simulation had finished, the final EC2 Instance wrote the final 3 files as comma-separated-value (CSV) files on the server. The author then downloaded the files to be used for further analysis. The total time to run the simulation was approximately 2 hours.

The general steps of the simulation are outlined below:

1. Using igraph, generate an empty graph of a given number of nodes.

2. Divide this size into roughly equal components, indexing from 1 to n, where lower indexes would correspond to higher levels in the network.

3. For each node n in each layer i, randomly choose a number of parents (up to a maximum of 10) from layers 1 to $i - 1$, and create the edges.

4. Calculate the number of structural holes, and randomly choose a number to convert them into full triads.

5. Take the resulting supply network and convert it into a risk graph by taking each node and edge and assigning it a unique node in the risk graph.

6. The edges in the risk graph were purely determined by the structure of the supply network.

7. Calculate the number of probability distributions that must be sampled for a given node/edge.

8. Sample from a Beta distribution with a given α and β

9. Assign the conditional probabilities in the risk graph to these sampled probabilities for each node/edge.

10. Calculate the risk measures using the graphs generated above and using bnlearn's simulation tools for probabilistic inference.

11. Use igraph to calculate the structural variables.

12. Take the results and store in a file.

13. Aggregate the files into three data sets (one data set for each structural level)

4.3 Description of Resulting Datasets

The resulting datasets were structure-level information for the 2-Ego Network, 1-Ego Network and Node/Arc Level. A total of 600 different networks were randomly sampled. This resulted in each data file comprising of an equal number of 322,000 observations. The variables in the Ego-2-Level were described above. The dataset comprised of 39 columns and 322,000 observations. The variables in the Ego-1-Level were also as described above, with the notable exception of the addition of the 2-Ego-Network Risks. The dataset comprised of 48 columns and 322,000 observations.

4.4 Empirical Models

In the design of the research, I have three measures the constructs of 2-Ego-Network Risk. Namely, the average 2-Ego-Network Risk, the standard deviation of 2-Ego-Network Risk and last the coefficient of variation of 2-Ego-Network Risk. Risk itself is decomposed into three sub-constructs, namely contribution, significance and velocity. I use the following notation to denote the variables:

1. C is Centrality
2. N is Connectivity
3. L is Clustering
4. F is Feedback
5. D is Distance
6. I is Risk Confidence
7. S is Risk Significance
8. V is Risk Velocity
9. E is Dummy Variable for Node(1)/Edge(0)

Given that I have 3 operationalizations each for 3 different risk constructs that use the same regressors, it is highly likely that the error terms in each model, across the models, will be correlated. As such, I leverage a Seemingly Unrelated Regression (SUR) model. In this case, because I have three different levels of analysis, of which are connected via 3 variables each (namely, the risk confidence, significance, and velocity, respectively), I estimate the SUR separately. Thus, there will be three systems of equations. The first will be the system of equations for the 2-Ego-Network risk model, the second for the 1-Ego-Network risk model, and the third for the local-network risk model. The models, given their length, are displayed on the following page.

$$\mu_c^{ego2} = \beta_{0,1} + \beta_{1,1}\alpha_i + \beta_{2,1}\beta_i + \beta_{3,1}C_i + \beta_{4,1}C_i^2 + \beta_{5,1}N_i + \beta_{6,1}N_i + \beta_{7,1}L_i + \beta_{8,1}L_i^2 + \beta_{9,1}F_i + \beta_{10,1}F_i^2 + \beta_{11,1}D_i + \beta_{12,1}D_i^2 + \epsilon_{i,1}^{ego2}$$

$$\mu_s^{ego2} = \beta_{0,2} + \beta_{1,2}\alpha_i + \beta_{2,2}\beta_i + \beta_{3,2}C_i + \beta_{4,2}C_i^2 + \beta_{5,2}N_i + \beta_{6,2}N_i + \beta_{7,2}L_i + \beta_{8,2}L_i^2 + \beta_{9,2}F_i + \beta_{10,2}F_i^2 + \beta_{11,2}D_i + \beta_{12,2}D_i^2 + \epsilon_{i,2}^{ego2}$$

$$\mu_v^{ego2} = \beta_{0,3} + \beta_{1,3}\alpha_i + \beta_{2,3}\beta_i + \beta_{3,3}C_i + \beta_{4,3}C_i^2 + \beta_{5,3}N_i + \beta_{6,3}N_i + \beta_{7,3}L_i + \beta_{8,3}L_i^2 + \beta_{9,3}F_i + \beta_{10,3}F_i^2 + \beta_{11,3}D_i + \beta_{12,3}D_i^2 + \epsilon_{i,3}^{ego2}$$

$$\sigma_c^{ego2} = \beta_{0,4} + \beta_{1,4}\alpha_i + \beta_{2,4}\beta_i + \beta_{3,4}C_i + \beta_{4,4}C_i^2 + \beta_{5,4}N_i + \beta_{6,4}N_i + \beta_{7,4}L_i + \beta_{8,4}L_i^2 + \beta_{9,4}F_i + \beta_{10,4}F_i^2 + \beta_{11,4}D_i + \beta_{12,4}D_i^2 + \epsilon_{i,4}^{ego2}$$

$$\sigma_s^{ego2} = \beta_{0,5} + \beta_{1,5}\alpha_i + \beta_{2,5}\beta_i + \beta_{3,5}C_i + \beta_{4,5}C_i^2 + \beta_{5,5}N_i + \beta_{6,5}N_i + \beta_{7,5}L_i + \beta_{8,5}L_i^2 + \beta_{9,5}F_i + \beta_{10,5}F_i^2 + \beta_{11,5}D_i + \beta_{12,5}D_i^2 + \epsilon_{i,5}^{ego2}$$

$$\sigma_v^{ego2} = \beta_{0,6} + \beta_{1,6}\alpha_i + \beta_{2,6}\beta_i + \beta_{3,6}C_i + \beta_{4,6}C_i^2 + \beta_{5,6}N_i + \beta_{6,6}N_i + \beta_{7,6}L_i + \beta_{8,6}L_i^2 + \beta_{9,6}F_i + \beta_{10,6}F_i^2 + \beta_{11,6}D_i + \beta_{12,6}D_i^2 + \epsilon_{i,6}^{ego2}$$

$$cv_c^{ego2} = \beta_{0,7} + \beta_{1,7}\alpha_i + \beta_{2,7}\beta_i + \beta_{3,7}C_i + \beta_{4,7}C_i^2 + \beta_{5,7}N_i + \beta_{6,7}N_i + \beta_{7,7}L_i + \beta_{8,7}L_i^2 + \beta_{9,7}F_i + \beta_{10,7}F_i^2 + \beta_{11,7}D_i + \beta_{12,7}D_i^2 + \epsilon_{i,7}^{ego2}$$

$$cv_s^{ego2} = \beta_{0,8} + \beta_{1,8}\alpha_i + \beta_{2,8}\beta_i + \beta_{3,8}C_i + \beta_{4,8}C_i^2 + \beta_{5,8}N_i + \beta_{6,8}N_i + \beta_{7,8}L_i + \beta_{8,8}L_i^2 + \beta_{9,8}F_i + \beta_{10,8}F_i^2 + \beta_{11,8}D_i + \beta_{12,8}D_i^2 + \epsilon_{i,8}^{ego2}$$

$$cv_v^{ego2} = \beta_{0,9} + \beta_{1,9}\alpha_i + \beta_{2,9}\beta_i + \beta_{3,9}C_i + \beta_{4,9}C_i^2 + \beta_{5,9}N_i + \beta_{6,9}N_i + \beta_{7,9}L_i + \beta_{8,9}L_i^2 + \beta_{9,9}F_i + \beta_{10,9}F_i^2 + \beta_{11,9}D_i + \beta_{12,9}D_i^2 + \epsilon_{i,9}^{ego2}$$

$$\mu_c^{ego1} = \beta_{0,1}^{ego1} + \beta_{1,1}^{ego1}\alpha_i + \beta_{2,1}^{ego1}\beta_i^{ego1} + \beta_{3,1}^{ego1}C_i + \beta_{4,1}^{ego1}C_i^2 + \beta_{5,1}^{ego1}N_i + \beta_{6,1}^{ego1}N_i + \beta_{7,1}^{ego1}L_i + \beta_{8,1}^{ego1}L_i^2 + \beta_{9,1}^{ego1}F_i + \beta_{10,1}^{ego1}F_i^2 + \beta_{11,1}^{ego1}D_i + \beta_{12,1}^{ego1}D_i^2 + \beta_{13,1}^{ego1}I_i + \beta_{14,1}^{ego1}S_i + \beta_{15,1}^{ego1}V_i + \epsilon_{i,1}^{ego1}$$

$$\mu_s^{ego1} = \beta_{0,2}^{ego1} + \beta_{1,2}^{ego1}\alpha_i + \beta_{2,2}^{ego1}\beta_i^{ego1} + \beta_{3,2}^{ego1}C_i + \beta_{4,2}^{ego1}C_i^2 + \beta_{5,2}^{ego1}N_i + \beta_{6,2}^{ego1}N_i + \beta_{7,2}^{ego1}L_i + \beta_{8,2}^{ego1}L_i^2 + \beta_{9,2}^{ego1}F_i + \beta_{10,2}^{ego1}F_i^2 + \beta_{11,2}^{ego1}D_i + \beta_{12,2}^{ego1}D_i^2 + \beta_{13,2}^{ego1}I_i + \beta_{14,2}^{ego1}S_i + \beta_{15,2}^{ego1}V_i + \epsilon_{i,2}^{ego1}$$

$$\mu_v^{ego1} = \beta_{0,3}^{ego1} + \beta_{1,3}^{ego1}\alpha_i + \beta_{2,3}^{ego1}\beta_i^{ego1} + \beta_{3,3}^{ego1}C_i + \beta_{4,3}^{ego1}C_i^2 + \beta_{5,3}^{ego1}N_i + \beta_{6,3}^{ego1}N_i + \beta_{7,3}^{ego1}L_i + \beta_{8,3}^{ego1}L_i^2 + \beta_{9,3}^{ego1}F_i + \beta_{10,3}^{ego1}F_i^2 + \beta_{11,3}^{ego1}D_i + \beta_{12,3}^{ego1}D_i^2 + \beta_{13,3}^{ego1}I_i + \beta_{14,3}^{ego1}S_i + \beta_{15,3}^{ego1}V_i + \epsilon_{i,3}^{ego1}$$

$$\sigma_c^{ego1} = \beta_{0,4}^{ego1} + \beta_{1,4}^{ego1}\alpha_i + \beta_{2,4}^{ego1}\beta_i^{ego1} + \beta_{3,4}^{ego1}C_i + \beta_{4,4}^{ego1}C_i^2 + \beta_{5,4}^{ego1}N_i + \beta_{6,4}^{ego1}N_i + \beta_{7,4}^{ego1}L_i + \beta_{8,4}^{ego1}L_i^2 + \beta_{9,4}^{ego1}F_i + \beta_{10,4}^{ego1}F_i^2 + \beta_{11,4}^{ego1}D_i + \beta_{12,4}^{ego1}D_i^2 + \beta_{13,4}^{ego1}I_i + \beta_{14,4}^{ego1}S_i + \beta_{15,4}^{ego1}V_i + \epsilon_{i,4}^{ego1}$$

$$\sigma_s^{ego1} = \beta_{0,5}^{ego1} + \beta_{1,5}^{ego1}\alpha_i + \beta_{2,5}^{ego1}\beta_i^{ego1} + \beta_{3,5}^{ego1}C_i + \beta_{4,5}^{ego1}C_i^2 + \beta_{5,5}^{ego1}N_i + \beta_{6,5}^{ego1}N_i + \beta_{7,5}^{ego1}L_i + \beta_{8,5}^{ego1}L_i^2 + \beta_{9,5}^{ego1}F_i + \beta_{10,5}^{ego1}F_i^2 + \beta_{11,5}^{ego1}D_i + \beta_{12,5}^{ego1}D_i^2 + \beta_{13,5}^{ego1}I_i + \beta_{14,5}^{ego1}S_i + \beta_{15,5}^{ego1}V_i + \epsilon_{i,5}^{ego1}$$

$$\sigma_v^{ego1} = \beta_{0,6}^{ego1} + \beta_{1,6}^{ego1}\alpha_i + \beta_{2,6}^{ego1}\beta_i^{ego1} + \beta_{3,6}^{ego1}C_i + \beta_{4,6}^{ego1}C_i^2 + \beta_{5,6}^{ego1}N_i + \beta_{6,6}^{ego1}N_i + \beta_{7,6}^{ego1}L_i + \beta_{8,6}^{ego1}L_i^2 + \beta_{9,6}^{ego1}F_i + \beta_{10,6}^{ego1}F_i^2 + \beta_{11,6}^{ego1}D_i + \beta_{12,6}^{ego1}D_i^2 + \beta_{13,6}^{ego1}I_i + \beta_{14,6}^{ego1}S_i + \beta_{15,6}^{ego1}V_i + \epsilon_{i,6}^{ego1}$$

$$cv_c^{ego1} = \beta_{0,7}^{ego1} + \beta_{1,7}^{ego1}\alpha_i + \beta_{2,7}^{ego1}\beta_i^{ego1} + \beta_{3,7}^{ego1}C_i + \beta_{4,7}^{ego1}C_i^2 + \beta_{5,7}^{ego1}N_i + \beta_{6,7}^{ego1}N_i + \beta_{7,7}^{ego1}L_i + \beta_{8,7}^{ego1}L_i^2 + \beta_{9,7}^{ego1}F_i + \beta_{10,7}^{ego1}F_i^2 + \beta_{11,7}^{ego1}D_i + \beta_{12,7}^{ego1}D_i^2 + \beta_{13,7}^{ego1}I_i + \beta_{14,7}^{ego1}S_i + \beta_{15,7}^{ego1}V_i + \epsilon_{i,7}^{ego1}$$

$$cv_s^{ego1} = \beta_{0,8}^{ego1} + \beta_{1,8}^{ego1}\alpha_i + \beta_{2,8}^{ego1}\beta_i^{ego1} + \beta_{3,8}^{ego1}C_i + \beta_{4,8}^{ego1}C_i^2 + \beta_{5,8}^{ego1}N_i + \beta_{6,8}^{ego1}N_i + \beta_{7,8}^{ego1}L_i + \beta_{8,8}^{ego1}L_i^2 + \beta_{9,8}^{ego1}F_i + \beta_{10,8}^{ego1}F_i^2 + \beta_{11,8}^{ego1}D_i + \beta_{12,8}^{ego1}D_i^2 + \beta_{13,8}^{ego1}I_i + \beta_{14,8}^{ego1}S_i + \beta_{15,8}^{ego1}V_i + \epsilon_{i,8}^{ego1}$$

$$cv_v^{ego1} = \beta_{0,9}^{ego1} + \beta_{1,9}^{ego1}\alpha_i + \beta_{2,9}^{ego1}\beta_i^{ego1} + \beta_{3,9}^{ego1}C_i + \beta_{4,9}^{ego1}C_i^2 + \beta_{5,9}^{ego1}N_i + \beta_{6,9}^{ego1}N_i + \beta_{7,9}^{ego1}L_i + \beta_{8,9}^{ego1}L_i^2 + \beta_{9,9}^{ego1}F_i + \beta_{10,9}^{ego1}F_i^2 + \beta_{11,9}^{ego1}D_i + \beta_{12,9}^{ego1}D_i^2 + \beta_{13,9}^{ego1}I_i + \beta_{14,9}^{ego1}S_i + \beta_{15,9}^{ego1}V_i + \epsilon_{i,9}^{ego1}$$

$$c^{local} = \beta_{0,7}^{loc} + \beta_{1,7}^{loc}\alpha_i + \beta_{2,7}^{loc}\beta_i^{loc} + \beta_{3,7}^{loc}C_i + \beta_{4,7}^{loc}C_i^2 + \beta_{5,7}^{loc}N_i + \beta_{6,7}^{loc}N_i + \beta_{7,7}^{loc}L_i + \beta_{8,7}^{loc}L_i^2 + \beta_{9,7}^{loc}F_i + \beta_{10,7}^{loc}F_i^2 + \beta_{11,7}^{loc}D_i + \beta_{12,7}^{loc}D_i^2 + \beta_{13,7}^{loc}I_i + \beta_{14,7}^{loc}S_i + \beta_{15,7}^{loc}V_i + \beta_{16,7}^{loc}E_i + \epsilon_{i,1}^{loc}$$

$$s^{local} = \beta_{0,8}^{loc} + \beta_{1,8}^{loc}\alpha_i + \beta_{2,8}^{loc}\beta_i^{loc} + \beta_{3,8}^{loc}C_i + \beta_{4,8}^{loc}C_i^2 + \beta_{5,8}^{loc}N_i + \beta_{6,8}^{loc}N_i + \beta_{7,8}^{loc}L_i + \beta_{8,8}^{loc}L_i^2 + \beta_{9,8}^{loc}F_i + \beta_{10,8}^{loc}F_i^2 + \beta_{11,8}^{loc}D_i + \beta_{12,8}^{loc}D_i^2 + \beta_{13,8}^{loc}I_i + \beta_{14,8}^{loc}S_i + \beta_{15,8}^{loc}V_i + \beta_{16,8}^{loc}E_i + \epsilon_{i,2}^{loc}$$

$$v^{lo} = \beta_{0,9}^{loc} + \beta_{1,9}^{loc}\alpha_i + \beta_{2,9}^{loc}\beta_i^{loc} + \beta_{3,9}^{loc}C_i + \beta_{4,9}^{loc}C_i^2 + \beta_{5,9}^{loc}N_i + \beta_{6,9}^{loc}N_i + \beta_{7,9}^{loc}L_i + \beta_{8,9}^{loc}L_i^2 + \beta_{9,9}^{loc}F_i + \beta_{10,9}^{loc}F_i^2 + \beta_{11,9}^{loc}D_i + \beta_{12,9}^{loc}D_i^2 + \beta_{13,9}^{loc}I_i + \beta_{14,9}^{loc}S_i + \beta_{15,9}^{loc}V_i + \beta_{16,9}^{loc}E_i + \epsilon_{i,3}^{loc}$$

Chapter 5

Analysis and Results

5.1 Descriptive Statistics

The raw network dataset that was generated by the simulation comprised of 600 distinct networks. The average number of nodes was 215.25, with a standard deviation of 10.066. The average number of arcs was 321.417, with a standard deviation of 63.484. We can notice, unfortunately, that the simulation resulted in networks that had extreme homogeneity within connectivity. The girth for all networks was 3, which indicates that every network had a triad. All other metrics had a fair level of heterogeneity. The entire summary statistics are shown in Table 5.1.

The three raw datasets were loaded into R and were initially cleaned by removing any duplicate entries that the simulation may have produced (it is possible that from a single network, duplicate ego-networks/2-ego networks corresponding to two different nodes/edges were recorded). After the duplicates were removed from the 2-Ego-Network dataset, the resulting number of observations was 193,147. The average size of the 2-Ego-Network comprised of 14.63 nodes with a standard deviation of 11.4. The average number of arcs was 26.1 with a standard deviation of 28.4. The summary for the rest of the variables is shown in Table 5.2. There are some notable outliers that must be addressed, namely in the risk variables. We can see that while the inner-quantile range is fairly tight, the outer quartiles contain data on a very different order of magnitude, indicating that further data cleansing is necessary.

Likewise, we notice that for the 1-Ego-Network data the number of observations, after removal of duplicates, amounted to a total sample size of 254,331. The average 1-Ego-Network size was 6.179 with a standard deviation of 3.7. The average number of arcs was 10.205 with a standard deviation of 11.601. We notice similar issues in the

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
eNNodes	600	215.250	10.066	200	206.8	223.2	236
eNArcs	600	321.417	63.484	210	268.8	370	478
eEdgeConnectivity	600	1.000	0.000	1	1	1	1
eNodeConnectivity	600	1.000	0.000	1	1	1	1
eDiameter	600	17.368	1.919	13	16	18	24
eRadius	600	8.935	0.996	7	8	9	12
eGirth	600	3.000	0.000	3	3	3	3
eCliqueNum	600	5.143	1.680	3	4	6	11
eNumOfLargestCliques	600	6.723	8.593	1	2	8	61
eAvCliqueSize	600	2.119	0.554	1.520	1.722	2.323	4.550
eNSH	600	660.505	150.013	381	531.8	769	1,228
eNTriads	600	158.537	123.190	5	59	234.2	588
eAvgConstraint	600	0.646	0.026	0.583	0.624	0.668	0.703
eConstraintCen	600	77.280	6.771	60.907	72.312	82.561	93.015
eDensity	600	0.014	0.003	0.009	0.012	0.016	0.022
eClusterCoeff	600	0.356	0.152	0.030	0.248	0.477	0.667
eClusterCoeffAv	600	0.324	0.145	0.016	0.215	0.439	0.622
eAvgShortestPathDis	600	7.399	0.610	5.820	6.976	7.815	9.323
eAvgBetweeness	600	686.120	79.019	508.112	630.537	741.615	944.684
eAvgCloseness	600	0.001	0.0001	0.0005	0.001	0.001	0.001
eAvgDegree	600	2.987	0.575	2.034	2.510	3.440	4.587
eAvgEV	600	0.066	0.013	0.028	0.057	0.073	0.121
eCenBetweeness	600	2,581,848.000	562,548.900	1,425,497	2,150,697.0	2,902,111.0	4,687,632
eCenEigen	600	201.153	10.359	179.266	192.683	209.408	229.430
eCenCloseness	600	19.813	2.307	13.214	18.153	21.332	26.668
eCenDegree	600	1,796.032	434.197	885	1,482.2	2,016.8	3,298
eAvgEccentricity	600	13.122	1.355	9.822	12.176	13.919	17.895
eEccentricityCen	600	914.870	160.686	503	801	1,002.2	1,558

Table 5.1: Summary Statistics of the Raw Network Data

risk measures as we noticed in the 2-Ego-Network data, which again would indicate that further data cleansing is needed. The results are shown in Table 5.3

The dataset for the nodes and edges comprised of the full 322,000 observations. There may be "duplicates", per se, however, they were not removed, as it is possible that one observation belonged to an edge while another belonged to a node, and hence are considered distinct, despite have the same exact characteristics. Recall that I seek to isolate the differences in the effect on risk between the nodes and the edges, and hence, removing duplicate information would prevent me from properly analyzing the two populations. The average value for the *isNode* variable was .401, which indicates that 40.1% of the observations were nodes while the remaining 59.9% were edges. Again we notice similar issues with the risk measures that were observed in the 1-Ego-Network and 2-Ego-Network data sets, hence indicating that further data cleansing is necessary. The correlation matrices are shown in Tables 5.5, 5.6, 5.7. Across all the data sets, we do see very high correlation between the number of nodes and the number of arcs, as

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
alpha	193,147	3.075	1.197	1	2	4	7
beta	193,147	5.512	2.867	1	3	8	10
eNNodes	193,147	14.763	11.353	3	6	20	83
eNArcs	193,147	26.056	28.438	2	7	34	243
eEdgeConnectivity	193,147	1.049	0.233	1	1	1	5
eNodeConnectivity	193,147	1.000	0.004	1	1	1	2
eDiameter	193,147	3.633	1.096	2	3	5	5
eRadius	193,147	2.069	0.693	1	2	3	3
eGirth	193,147	2.541	1.080	0	3	3	3
eCliqueNum	193,147	3.863	1.531	2	3	5	11
eNumOfLargestCliques	193,147	2.732	2.472	1	1	3	37
eAvCliqueSize	193,147	2.025	0.601	1.400	1.636	2.244	5.499
eNSH	193,147	50.253	71.446	1	6	63	877
eNTriads	193,147	21.150	40.042	0	1	21	443
eAvgConstraint	193,147	0.720	0.095	0.374	0.651	0.791	0.924
eConstraintCen	193,147	4.815	4.878	0.284	1.250	6.660	36.388
eDensity	193,147	0.326	0.181	0.036	0.176	0.467	0.848
eClusterCoeff	193,147	0.412	0.238	0.000	0.263	0.600	0.980
eClusterCoeffAv	193,147	0.466	0.261	0.000	0.303	0.666	0.977
eAvgShortestPathDis	193,147	2.152	0.558	1.152	1.667	2.607	3.739
eAvgBetweeness	193,147	10.512	12.092	0.333	1.667	15.050	92.169
eAvgCloseness	193,147	0.089	0.099	0.004	0.021	0.120	0.389
eAvgDegree	193,147	2.929	1.147	1.333	2.000	3.600	9.846
eAvgEV	193,147	0.518	0.172	0.143	0.375	0.652	0.917
eCenBetweeness	193,147	2,248.995	5,700.472	2.000	32.000	1,521.000	113,007.400
eCenEigen	193,147	8.740	8.902	0.586	2.090	12.451	66.816
eCenCloseness	193,147	3.000	2.198	0.500	1.333	4.009	15.873
eCenDegree	193,147	79.103	105.544	2	10	103	1,134
eAvgEccentricity	193,147	3.147	0.931	1.667	2.500	4.000	4.750
eEccentricityCen	193,147	8.205	8.519	1	2	11	60
avg_elrcf	193,147	10.728	11.100	0.128	3.132	14.592	149.428
avg_elrsf	193,147	6.273	3.889	0.257	3.584	7.890	47.390
avg_elrv	193,147	4.815	7.770	−0.977	−0.297	7.759	120.174
sd_elrcf	193,147	19.372	21.275	0.056	3.723	28.856	193.074
sd_elrsf	193,147	1.795	1.460	0.072	0.862	2.253	21.200
sd_elrv	193,147	14.236	19.170	0.018	1.032	22.776	174.067
cv_elrcf	193,147	1.576	0.586	0.200	1.126	1.976	4.776
cv_elrsf	193,147	0.321	0.192	0.015	0.165	0.453	1.467
cv_elrv	193,147	5.053	3,057.665	−296,747.800	−0.384	3.590	1,237,882.000

Table 5.2: Summary Statistics for The 2-Ego-Network Original Data-Set

should be expected. In addition, we notice that the diameter of the network is highly correlated with the radius (which is not necessarily obvious, despite the names) and the average shortest path distance. The number of structural holes is correlated very high with the constraint (as expected) as well as with the number of triads. We also notice fairly high correlations between the various centrality measures (closeness, eigenvector, betweenness, and degree), which again was expected.

However, we also notice that the dependent variables across all the data sets seem to be highly correlated. This may be an indication that the error terms across the models, if we were to run them individually, would result in cross-equation correlation

and possible endogeneity. This provides enough justification for the model specification and estimation approach that I will explain below.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
avg_elrcf_2ego	254,331	10.682	11.058	0.128	3.193	14.440	149.428
avg_elrsf_2ego	254,331	6.364	3.933	0.257	3.645	8.007	47.390
avg_elrv_2ego	254,331	4.741	7.725	−0.977	−0.288	7.547	120.174
sd_elrcf_2ego	254,331	19.354	21.241	0.056	3.869	28.605	193.074
sd_elrsf_2ego	254,331	1.840	1.478	0.072	0.894	2.312	21.200
sd_elrv_2ego	254,331	14.092	19.109	0.018	1.063	22.284	174.067
cv_elrcf_2ego	254,331	1.593	0.580	0.200	1.149	1.981	4.776
cv_elrsf_2ego	254,331	0.323	0.191	0.015	0.169	0.453	1.467
cv_elrv_2ego	254,331	3.392	2,717.067	−296,747.800	−0.442	3.637	1,237,882.000
alpha	254,331	3.065	1.192	1	2	4	7
beta	254,331	5.527	2.864	1	3	8	10
eNNodes	254,331	6.179	3.700	2	3	8	29
eNArcs	254,331	10.205	11.601	1	3	13	129
eEdgeConnectivity	254,331	1.377	0.748	1	1	2	10
eNodeConnectivity	254,331	1.187	0.597	1	1	1	10
eDiameter	254,331	2.148	0.770	1	2	3	3
eRadius	254,331	1.382	0.486	1	1	2	2
eGirth	254,331	2.170	1.342	0	0	3	3
eCliqueNum	254,331	3.412	1.394	2	2	4	11
eNumOfLargestCliques	254,331	1.859	1.452	1	1	2	31
eAvCliqueSize	254,331	1.908	0.592	1.333	1.444	2.129	5.503
eNSH	254,331	12.311	20.524	0	1	15	485
eNTriads	254,331	8.989	20.901	0	0	8	287
eAvgConstraint	254,331	0.812	0.159	0.242	0.720	0.906	1.125
eConstraintCen	254,331	1.327	1.581	0.000	0.099	1.838	16.174
eDensity	254,331	0.630	0.243	0.111	0.429	0.833	1.000
eClusterCoeff	254,331	0.465	0.339	0.000	0.000	0.727	1.000
eClusterCoeffAv	254,331	0.508	0.357	0.000	0.000	0.810	1.000
eAvgShortestPathDis	254,331	1.434	0.316	1.000	1.167	1.682	2.287
eAvgBetweeness	254,331	1.554	1.731	0.000	0.250	2.429	12.667
eAvgCloseness	254,331	0.302	0.312	0.020	0.086	0.389	1.000
eAvgDegree	254,331	2.610	1.397	1.000	1.600	3.250	11.176
eAvgEV	254,331	0.753	0.173	0.285	0.612	0.890	1.000
eCenBetweeness	254,331	94.293	207.940	0.000	1.000	87.000	4,100.010
eCenEigen	254,331	2.030	2.136	0.000	0.438	3.124	14.982
eCenCloseness	254,331	1.357	1.187	0.000	0.500	2.076	7.787
eCenDegree	254,331	17.980	26.511	0	2	24	322
eAvgEccentricity	254,331	1.898	0.623	1.000	1.500	2.538	2.905
eEccentricityCen	254,331	1.860	1.907	0	1	3	14
avg_elrcf	254,331	9.136	13.292	0.056	1.802	10.271	218.835
avg_elrsf	254,331	6.597	4.263	0.169	3.651	8.345	67.955
avg_elrv	254,331	3.721	9.597	−0.991	−0.693	2.707	195.706
sd_elrcf	254,331	13.523	21.050	0.023	1.552	15.218	225.270
sd_elrsf	254,331	1.330	1.282	0.015	0.564	1.632	22.486
sd_elrv	254,331	9.063	18.535	0.008	0.349	6.582	225.798
cv_elrcf	254,331	1.208	0.523	0.023	0.822	1.515	4.776
cv_elrsf	254,331	0.244	0.205	0.005	0.100	0.318	1.485
cv_elrv	254,331	1.653	2,680.906	−296,747.800	−0.582	2.589	1,237,882.000

Table 5.3: Summary Statistics for The 1-Ego-Network Original Data-Set

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
alpha	322,000	3.063	1.191	1	2	4	7
beta	322,000	5.525	2.864	1	3	8	10
avg_elrcf_ego	322,000	8.879	13.040	0.056	1.823	9.811	218.835
avg_elrsf_ego	322,000	6.652	4.288	0.169	3.685	8.417	67.955
avg_elrv_ego	322,000	3.547	9.436	−0.991	−0.689	2.424	195.706
sd_elrcf_ego	322,000	13.208	20.840	0.023	1.598	14.448	225.270
sd_elrsf_ego	322,000	1.338	1.269	0.015	0.579	1.640	22.486
sd_elrv_ego	322,000	8.763	18.329	0.008	0.359	6.047	225.798
cv_elrcf_ego	322,000	1.217	0.521	0.023	0.835	1.517	4.776
cv_elrsf_ego	322,000	0.243	0.202	0.005	0.102	0.312	1.485
cv_elrv_ego	322,000	4.420	3,264.272	−296,747.800	−0.640	2.606	1,237,882.000
isNode	322,000	0.401	0.490	0	0	1	1
nBetweenness	322,000	818.526	2,151.709	0.000	0.000	618.000	51,751.500
nCloseness	322,000	0.001	0.0002	0.0002	0.0004	0.001	0.001
nEigen	322,000	0.082	0.184	0.000	0.0003	0.051	1.000
nDegree	322,000	5.432	4.352	1	2	7	36
nEccentricity	322,000	12.626	2.249	6	11	14	24
nMinCenterDistance	322,000	3.918	1.849	0	3	5	12
nMaxCenterDistance	322,000	4.493	1.817	0	3	6	12
nAvgSPathDis	322,000	6.858	1.638	3.042	5.649	7.906	16.348
nDensity	322,000	0.018	0.012	0.002	0.009	0.025	0.098
nClusterCoeff	322,000	0.482	0.328	0.000	0.267	0.667	1.000
nCoreness	322,000	3.803	2.877	1	2	5	19
nTriads	322,000	11.694	20.525	0	1	13	316
nSH	322,000	9.811	18.143	0	0	11	314
nEgoConnectivity	322,000	0.858	0.719	0	0	1	9
elrsf	322,000	6.775	4.683	0.019	3.549	8.736	78.592
elrcf	322,000	6.772	22.569	0.006	0.641	4.044	529.657
elrv	322,000	2.322	18.147	−1.000	−0.938	−0.085	528.897

Table 5.4: Summary Statistics for The Node/Edge Original Data-Set

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
1.alpha																																						
2.beta	0.00																																					
3.eNNodes	-0.02	0.01																																				
4.eNArcs	-0.04	0.02	0.94																																			
5.eEdgeConnectivity	-0.02	0.01	-0.05	0.00																																		
6.eNodeConnectivity	0.00	0.00	0.00	0.00	0.02																																	
7.eDiameter	-0.01	0.01	0.71	0.58	-0.16	-0.01																																
8.eRadius	-0.01	0.01	0.68	0.55	-0.14	-0.01	0.95																															
9.eGirth	-0.03	0.04	0.37	0.33	0.09	0.00	0.36	0.34																														
10.eCliqueNum	-0.07	0.04	0.65	0.78	0.11	0.00	0.40	0.38	0.52																													
11.eNumOfLargestCliques	0.02	-0.03	0.27	0.19	-0.04	0.00	0.25	0.24	-0.19	-0.17																												
12.eAvCliqueSize	-0.07	0.04	0.60	0.77	0.13	0.00	0.33	0.32	0.42	0.97	-0.05																											
13.eNSH	-0.03	0.00	0.93	0.95	-0.04	0.00	0.54	0.51	0.27	0.63	0.27	0.61																										
14.eNTriads	-0.06	0.01	0.74	0.91	0.03	0.00	0.37	0.36	0.22	0.81	0.11	0.85	0.83																									
15.eAvgConstraint	0.04	-0.03	-0.69	-0.75	-0.08	0.00	-0.62	-0.59	-0.35	-0.77	-0.17	-0.78	-0.64	-0.67																								
16.eConstraintCen	-0.03	0.02	0.97	0.98	-0.09	0.00	0.65	0.62	0.33	0.72	0.23	0.69	0.95	0.84	-0.76																							
17.eDensity	0.00	0.01	-0.76	-0.61	0.11	0.00	-0.87	-0.82	-0.38	-0.37	-0.32	-0.30	-0.59	-0.37	0.60	-0.67																						
18.eClusterCoeff	-0.07	0.06	0.27	0.40	0.20	0.00	0.12	0.12	0.74	0.76	-0.25	0.72	0.25	0.43	-0.47	0.34	-0.02																					
19.eClusterCoeffAv	-0.06	0.06	0.37	0.47	0.31	0.00	0.17	0.16	0.76	0.72	-0.15	0.68	0.35	0.44	-0.50	0.41	-0.16	0.92																				
20.eAvgShortestPathDis	0.00	0.00	0.81	0.64	-0.13	0.00	0.93	0.87	0.36	0.37	0.31	0.30	0.62	0.38	-0.57	0.72	-0.94	0.05	0.15																			
21.eAvgBetweeness	-0.02	0.01	0.98	0.88	-0.08	0.00	0.71	0.68	0.31	0.53	0.28	0.47	0.88	0.65	-0.58	0.92	-0.74	0.17	0.27	0.84																		
22.eAvgCloseness	0.01	-0.01	-0.64	-0.54	-0.02	0.00	-0.78	-0.74	-0.63	-0.53	-0.21	-0.46	-0.48	-0.36	0.69	-0.58	0.88	-0.38	-0.46	-0.79	-0.57																	
23.eAvgDegree	-0.06	0.04	0.68	0.81	0.16	0.00	0.42	0.40	0.53	0.94	0.04	0.95	0.67	0.80	-0.88	0.76	-0.42	0.75	0.76	0.41	0.55	-0.60																
24.eAvgEV	0.00	0.00	-0.83	-0.70	0.07	0.00	-0.83	-0.78	-0.46	-0.48	-0.28	-0.41	-0.67	-0.46	0.65	-0.76	0.95	-0.15	-0.28	-0.92	-0.80	0.84	-0.53															
25.eCenBetweeness	-0.02	0.00	0.82	0.78	-0.05	0.00	0.41	0.41	0.16	0.41	0.22	0.38	0.83	0.64	-0.38	0.79	-0.44	0.12	0.19	0.53	0.85	-0.31	0.41	-0.51														
26.eCenEigen	-0.02	0.01	0.99	0.92	-0.07	0.00	0.70	0.67	0.33	0.59	0.27	0.54	0.92	0.71	-0.64	0.95	-0.75	0.21	0.32	0.81	0.99	-0.60	0.62	-0.83	0.84													
27.eCenCloseness	-0.02	0.01	0.97	0.91	-0.02	0.00	0.62	0.60	0.38	0.62	0.27	0.58	0.91	0.71	-0.67	0.94	-0.73	0.26	0.37	0.75	0.94	-0.62	0.67	-0.83	0.81	0.97												
28.eCenDegree	-0.02	0.00	0.95	0.94	-0.05	0.00	0.57	0.55	0.28	0.63	0.26	0.61	0.97	0.81	-0.63	0.95	-0.63	0.23	0.33	0.66	0.91	-0.50	0.65	-0.72	0.85	0.95	0.94											
29.eAvgEccentricity	0.00	0.01	0.74	0.60	-0.15	-0.01	0.99	0.93	0.39	0.42	0.27	0.34	0.55	0.38	-0.63	0.67	-0.90	0.13	0.19	0.96	0.74	-0.80	0.44	-0.87	0.44	0.73	0.66	0.59										
30.eEccentricityCen	-0.03	0.01	0.88	0.84	-0.09	0.00	0.75	0.74	0.30	0.58	0.22	0.54	0.83	0.69	-0.62	0.87	-0.68	0.23	0.30	0.74	0.85	-0.56	0.59	-0.73	0.68	0.87	0.83	0.84	0.71									
31.avg_elrcf	0.18	-0.33	0.44	0.44	0.02	0.01	0.30	0.28	0.25	0.41	0.13	0.40	0.40	0.37	-0.45	0.44	-0.37	0.23	0.28	0.35	0.38	-0.38	0.46	-0.43	0.25	0.42	0.45	0.41	0.33	0.34								
32.avg_elrsf	0.28	-0.54	-0.09	-0.02	0.06	0.00	-0.09	-0.08	0.00	0.07	-0.08	0.08	-0.04	0.05	0.00	-0.05	0.17	0.14	0.09	-0.15	-0.11	0.11	0.05	0.17	-0.05	-0.10	-0.11	-0.06	-0.11	-0.04	0.05							
33.avg_elrv	0.08	-0.13	0.41	0.39	0.00	0.02	0.28	0.26	0.22	0.33	0.14	0.32	0.37	0.31	-0.39	0.40	-0.36	0.16	0.21	0.35	0.37	-0.35	0.38	-0.42	0.24	0.40	0.43	0.39	0.32	0.31	0.90	-0.22						
34.sd_elrcf	0.16	-0.31	0.50	0.49	0.00	0.01	0.36	0.33	0.26	0.42	0.15	0.41	0.46	0.40	-0.47	0.50	-0.43	0.22	0.27	0.42	0.45	-0.41	0.47	-0.48	0.31	0.48	0.51	0.47	0.39	0.40	0.95	-0.01	0.90					
35.sd_elrsf	0.18	-0.42	0.52	0.61	0.05	0.00	0.34	0.32	0.29	0.61	0.12	0.63	0.58	0.61	-0.60	0.58	-0.33	0.41	0.41	0.33	0.44	-0.39	0.65	-0.39	0.35	0.49	0.51	0.54	0.35	0.48	0.65	0.43	0.43	0.65				
36.sd_elrv	0.10	-0.19	0.50	0.48	-0.01	0.01	0.35	0.32	0.23	0.37	0.17	0.36	0.46	0.38	-0.43	0.49	-0.42	0.16	0.22	0.42	0.47	-0.38	0.42	-0.48	0.33	0.49	0.51	0.48	0.38	0.40	0.88	-0.17	0.95	0.95	0.52			
37.cv_elrcf	-0.01	0.00	0.52	0.44	-0.01	0.00	0.48	0.45	0.33	0.35	0.17	0.31	0.42	0.30	-0.43	0.47	-0.56	0.17	0.24	0.54	0.49	-0.53	0.40	-0.59	0.31	0.51	0.53	0.45	0.51	0.45	0.38	-0.28	0.44	0.56	0.27	0.56		
38.cv_elrsf	-0.03	-0.01	0.67	0.69	0.01	0.01	0.48	0.45	0.37	0.62	0.17	0.62	0.65	0.59	-0.69	0.69	-0.53	0.36	0.42	0.52	0.60	-0.55	0.69	-0.60	0.43	0.65	0.68	0.65	0.51	0.57	0.66	-0.29	0.73	0.70	0.59	0.75	0.58	
39.cv_elrv	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 5.5: The 2-Ego-Network Correlation Matrix. Gray cells indicate $p < 0.05$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
1.avg_elrcf_2ego																																															
2.avg_elrsf_2ego	0.05																																														
3.avg_elrv_2ego	0.90	-0.22																																													
4.sd_elrcf_2ego	0.95	-0.01	0.90																																												
5.sd_elrsf_2ego	0.65	0.44	0.43	0.64																																											
6.sd_elrv_2ego	0.88	-0.17	0.96	0.94	0.52																																										
7.cv_elrcf_2ego	0.36	-0.28	0.43	0.55	0.25	0.55																																									
8.cv_elrsf_2ego	0.66	-0.29	0.73	0.70	0.58	0.75	0.57																																								
9.cv_elrv_2ego	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																																							
10.alpha	0.17	0.28	0.08	0.16	0.18	0.10	-0.02	-0.03	0.00																																						
11.beta	-0.33	-0.54	-0.13	-0.31	-0.43	-0.19	0.00	-0.01	0.00	0.01																																					
12.eNNodes	0.34	-0.04	0.31	0.41	0.47	0.41	0.48	0.58	0.00	-0.03	0.01																																				
13.eNArcs	0.33	0.02	0.28	0.38	0.54	0.37	0.38	0.57	0.00	-0.04	0.02	0.91																																			
14.eEdgeConnectivity	0.13	0.07	0.09	0.13	0.23	0.11	0.10	0.20	0.00	-0.03	0.02	0.23	0.36																																		
15.eNodeConnectivity	0.00	0.06	-0.01	-0.01	0.05	-0.02	-0.04	0.01	0.00	-0.01	0.01	-0.13	-0.02	0.72																																	
16.eDiameter	0.23	-0.06	0.21	0.28	0.28	0.27	0.39	0.38	0.00	-0.01	0.01	0.72	0.51	-0.15	-0.33																																
17.eRadius	0.19	-0.05	0.18	0.23	0.23	0.23	0.33	0.32	0.00	-0.01	0.01	0.66	0.48	-0.13	-0.25	0.87																															
18.eGirth	0.22	0.03	0.19	0.25	0.32	0.23	0.33	0.37	0.00	-0.03	0.04	0.54	0.44	0.31	0.19	0.43	0.32																														
19.eCliqueNum	0.31	0.07	0.25	0.35	0.53	0.32	0.35	0.53	0.00	-0.06	0.03	0.75	0.86	0.44	0.16	0.41	0.34	0.63																													
20.eNumOfLargestCliques	0.13	-0.06	0.14	0.17	0.17	0.18	0.21	0.22	0.00	0.01	-0.02	0.40	0.32	0.03	-0.09	0.39	0.33	-0.03	0.00																												
21.eAvCliqueSize	0.31	0.07	0.25	0.35	0.55	0.32	0.34	0.54	0.00	-0.06	0.03	0.75	0.88	0.47	0.18	0.39	0.32	0.56	0.98	0.12																											
22.eNSH	0.28	-0.02	0.26	0.34	0.46	0.35	0.36	0.51	0.00	-0.02	0.00	0.88	0.89	0.19	-0.13	0.49	0.49	0.33	0.59	0.45	0.61																										
23.eNTriads	0.26	0.06	0.21	0.30	0.51	0.28	0.25	0.46	0.00	-0.05	0.01	0.69	0.91	0.36	0.04	0.29	0.28	0.27	0.80	0.23	0.86	0.74																									
24.eAvgConstraint	-0.32	0.02	-0.29	-0.37	-0.48	-0.36	-0.42	-0.55	0.00	0.03	-0.02	-0.84	-0.81	-0.21	0.12	-0.75	-0.57	-0.39	-0.72	-0.41	-0.75	-0.69	-0.63																								
25.eConstraintCen	0.29	-0.03	0.27	0.35	0.45	0.35	0.39	0.53	0.00	-0.03	0.01	0.86	0.82	-0.10	-0.25	0.66	0.65	0.38	0.67	0.38	0.68	0.81	0.67	-0.77																							
26.eDensity	-0.23	0.11	-0.23	-0.29	-0.22	-0.29	-0.42	-0.37	0.00	-0.01	0.00	-0.73	-0.46	0.17	0.41	-0.91	-0.71	-0.40	-0.30	-0.44	-0.28	-0.51	-0.21	0.70	-0.61																						
27.eClusterCoeff	0.18	0.10	0.12	0.19	0.33	0.15	0.20	0.31	0.00	-0.05	0.04	0.36	0.41	0.52	0.44	0.12	0.08	0.85	0.69	-0.13	0.65	0.20	0.34	-0.23	0.24	-0.01																					
28.eClusterCoeffAv	0.21	0.08	0.16	0.23	0.35	0.20	0.26	0.36	0.00	-0.05	0.04	0.47	0.47	0.56	0.39	0.19	0.13	0.88	0.69	-0.04	0.65	0.30	0.35	-0.31	0.30	-0.13	0.96																				
29.eAvgShortestPathDis	0.22	-0.11	0.23	0.28	0.21	0.29	0.42	0.36	0.00	0.01	0.00	0.74	0.46	-0.16	-0.38	0.92	0.80	0.38	0.29	0.44	0.26	0.51	0.20	-0.66	0.61	-0.98	0.00	0.12																			
30.eAvgBetweeness	0.28	-0.09	0.27	0.35	0.34	0.36	0.45	0.47	0.00	-0.01	0.00	0.93	0.75	0.05	-0.26	0.75	0.75	0.40	0.50	0.42	0.48	0.81	0.49	-0.70	0.78	-0.79	0.13	0.27	0.84																		
31.eAvgCloseness	-0.27	0.04	-0.24	-0.31	-0.34	-0.29	-0.43	-0.44	0.00	0.02	-0.02	-0.70	-0.52	-0.16	0.03	-0.81	-0.54	-0.71	-0.56	-0.34	-0.53	-0.44	-0.31	0.73	-0.56	0.83	-0.49	-0.55	-0.78	-0.61																	
32.eAvgDegree	0.34	0.05	0.28	0.38	0.56	0.36	0.39	0.58	0.00	-0.05	0.03	0.83	0.92	0.49	0.18	0.48	0.39	0.62	0.95	0.24	0.97	0.71	0.83	-0.82	0.73	-0.39	0.65	0.68	0.37	0.58	-0.63																
33.eAvgEV	-0.26	0.12	-0.26	-0.32	-0.25	-0.32	-0.45	-0.41	0.00	0.00	0.00	-0.79	-0.53	0.10	0.38	-0.86	-0.67	-0.44	-0.36	-0.43	-0.33	-0.59	-0.26	0.71	-0.65	0.98	-0.05	-0.20	-0.95	-0.84	0.81	-0.44															
34.eCenBetweeness	0.23	-0.03	0.22	0.29	0.35	0.30	0.32	0.41	0.00	-0.02	0.00	0.82	0.82	0.19	-0.14	0.42	0.44	0.26	0.53	0.33	0.54	0.90	0.69	-0.56	0.69	-0.46	0.15	0.25	0.49	0.83	-0.35	0.59	-0.54														
35.eCenEigen	0.30	-0.09	0.29	0.37	0.36	0.38	0.46	0.49	0.00	-0.01	0.00	0.95	0.78	0.08	-0.26	0.72	0.68	0.43	0.54	0.42	0.52	0.85	0.52	-0.73	0.79	-0.80	0.15	0.30	0.81	0.98	-0.63	0.62	-0.87	0.84													
36.eCenCloseness	0.32	-0.08	0.30	0.38	0.39	0.39	0.47	0.52	0.00	-0.01	0.01	0.93	0.79	0.08	-0.30	0.70	0.56	0.47	0.60	0.41	0.59	0.82	0.55	-0.79	0.79	-0.81	0.20	0.34	0.78	0.89	-0.71	0.68	-0.89	0.77	0.95												
37.eCenDegree	0.29	-0.06	0.27	0.35	0.40	0.36	0.40	0.49	0.00	-0.02	0.00	0.92	0.86	0.16	-0.19	0.53	0.52	0.36	0.60	0.38	0.60	0.94	0.68	-0.69	0.80	-0.60	0.18	0.31	0.60	0.89	-0.49	0.68	-0.69	0.94	0.94	0.91											
38.eAvgEccentricity	0.24	-0.08	0.23	0.29	0.28	0.29	0.41	0.40	0.00	0.00	0.01	0.75	0.52	-0.15	-0.36	0.98	0.87	0.44	0.40	0.40	0.38	0.51	0.28	-0.74	0.67	-0.94	0.11	0.19	0.96	0.80	-0.81	0.47	-0.90	0.46	0.77	0.75	0.58										
39.eEccentricityCen	0.26	-0.01	0.23	0.31	0.40	0.30	0.36	0.45	0.00	-0.03	0.02	0.79	0.75	0.06	-0.15	0.78	0.77	0.40	0.62	0.34	0.63	0.69	0.60	-0.75	0.78	-0.62	0.26	0.31	0.63	0.71	-0.60	0.69	-0.62	0.58	0.71	0.66	0.67	0.72									
40.avg_elrcf	0.70	0.02	0.61	0.68	0.53	0.65	0.27	0.56	0.00	0.12	-0.24	0.51	0.50	0.17	-0.01	0.30	0.26	0.27	0.42	0.23	0.43	0.48	0.40	-0.45	0.45	-0.32	0.19	0.25	0.31	0.44	-0.33	0.47	-0.37	0.41	0.47	0.48	0.48	0.32	0.36								
41.avg_elrsf	0.05	0.97	-0.20	-0.01	0.44	-0.17	-0.29	-0.28	0.00	0.27	-0.52	-0.12	-0.06	0.04	0.06	-0.11	-0.09	-0.01	0.00	-0.09	0.01	-0.09	0.00	0.09	-0.09	0.16	0.07	0.04	-0.16	-0.16	0.09	-0.02	0.17	-0.10	-0.16	-0.15	-0.13	-0.13	-0.06	-0.09							
42.avg_elrv	0.57	-0.16	0.61	0.57	0.32	0.61	0.27	0.57	0.00	0.05	-0.08	0.45	0.43	0.13	-0.01	0.25	0.22	0.21	0.33	0.21	0.34	0.43	0.33	-0.38	0.39	-0.29	0.13	0.19	0.28	0.40	-0.27	0.39	-0.33	0.38	0.43	0.43	0.43	0.27	0.31	0.92	-0.25						
43.sd_elrcf	0.74	-0.03	0.68	0.73	0.52	0.71	0.31	0.61	0.00	0.11	-0.22	0.54	0.52	0.17	-0.02	0.31	0.28	0.27	0.43	0.24	0.43	0.50	0.42	-0.46	0.46	-0.34	0.19	0.25	0.33	0.47	-0.33	0.48	-0.38	0.44	0.50	0.50	0.51	0.33	0.38	0.95	-0.13	0.90					
44.sd_elrsf	0.52	0.34	0.34	0.54	0.83	0.45	0.27	0.51	0.00	0.15	-0.36	0.61	0.71	0.27	0.06	0.36	0.30	0.37	0.66	0.26	0.69	0.61	0.67	-0.61	0.60	-0.29	0.38	0.40	0.27	0.44	-0.42	0.72	-0.32	0.47	0.47	0.51	0.53	0.35	0.52	0.60	0.28	0.42	0.60				
45.sd_elrv	0.60	-0.15	0.65	0.62	0.37	0.67	0.30	0.60	0.00	0.06	-0.12	0.49	0.47	0.14	-0.02	0.27	0.25	0.22	0.36	0.23	0.37	0.48	0.38	-0.41	0.43	-0.31	0.14	0.20	0.30	0.44	-0.29	0.42	-0.35	0.43	0.47	0.47	0.48	0.30	0.34	0.89	-0.23	0.96	0.94	0.49			
46.cv_elrcf	0.47	-0.23	0.50	0.50	0.24	0.50	0.44	0.54	0.00	-0.01	0.01	0.48	0.40	0.12	-0.02	0.37	0.32	0.32	0.36	0.20	0.35	0.37	0.27	-0.42	0.39	-0.39	0.20	0.27	0.39	0.45	-0.40	0.40	-0.43	0.33	0.46	0.47	0.41	0.39	0.36	0.36	-0.26	0.37	0.53	0.30	0.48		
47.cv_elrsf	0.48	-0.24	0.53	0.54	0.44	0.59	0.49	0.79	0.00	-0.02	0.00	0.72	0.73	0.25	0.03	0.44	0.38	0.42	0.65	0.29	0.67	0.67	0.61	-0.69	0.67	-0.42	0.35	0.40	0.41	0.59	-0.50	0.73	-0.47	0.55	0.62	0.64	0.64	0.46	0.56	0.69	-0.32	0.73	0.73	0.65	0.75	0.54	
48.cv_elrv	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 5.6: The 1-Ego-Network Correlation Matrix. Gray cells indicate $p < 0.05$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1.alpha																												
2.beta	0.01																											
3.avg_elrcf_ego	0.12	-0.23																										
4.avg_elrsf_ego	0.27	-0.52	-0.08																									
5.avg_elrv_ego	0.05	-0.08	0.92	-0.25																								
6.sd_elrcf_ego	0.11	-0.22	0.95	-0.12	0.90																							
7.sd_elrsf_ego	0.15	-0.36	0.60	0.29	0.42	0.60																						
8.sd_elrv_ego	0.06	-0.11	0.89	-0.23	0.96	0.94	0.48																					
9.cv_elrcf_ego	-0.01	0.00	0.35	-0.26	0.36	0.52	0.29	0.47																				
10.cv_elrsf_ego	-0.02	0.00	0.69	-0.32	0.73	0.73	0.64	0.75	0.53																			
11.cv_elrv_ego	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																		
12.isNode	0.01	-0.01	-0.16	0.04	-0.13	-0.18	-0.24	-0.16	-0.26	-0.28	0.00																	
13.nBetweenness	-0.01	0.00	0.33	-0.12	0.32	0.31	0.14	0.31	0.15	0.30	0.00	-0.05																
14.nCloseness	0.05	-0.02	0.18	-0.15	0.20	0.17	-0.04	0.17	0.10	0.11	0.00	0.46	0.26															
15.nEigen	-0.01	0.00	0.25	-0.07	0.25	0.25	0.19	0.25	0.16	0.29	0.00	-0.07	0.33	0.41														
16.nDegree	-0.03	0.02	0.27	0.00	0.23	0.27	0.33	0.24	0.24	0.36	0.00	-0.46	0.39	0.08	0.62													
17.nEccentricity	-0.01	0.00	-0.26	0.08	-0.24	-0.27	-0.21	-0.24	-0.24	-0.30	0.00	0.18	-0.26	-0.53	-0.40	-0.47												
18.nMinCenterDistance	0.00	0.01	-0.29	0.12	-0.27	-0.29	-0.19	-0.27	-0.25	-0.31	0.00	0.12	-0.33	-0.59	-0.47	-0.49	0.89											
19.nMaxCenterDistance	0.00	0.00	-0.29	0.12	-0.27	-0.29	-0.19	-0.27	-0.25	-0.31	0.00	0.11	-0.33	-0.61	-0.47	-0.49	0.89	0.95										
20.nAvgSPathDis	0.01	-0.01	-0.33	0.10	-0.29	-0.33	-0.27	-0.30	-0.31	-0.38	0.00	0.27	-0.33	-0.55	-0.50	-0.63	0.86	0.88	0.88									
21.nDensity	-0.01	0.01	0.29	-0.05	0.26	0.28	0.29	0.26	0.24	0.37	0.00	-0.28	0.44	0.35	0.68	0.92	-0.49	-0.54	-0.54	-0.64								
22.nClusterCoeff	-0.02	0.02	0.09	0.03	0.07	0.10	0.20	0.07	0.16	0.20	0.00	-0.39	-0.07	-0.10	0.06	0.21	-0.15	-0.12	-0.12	-0.21	0.19							
23.nCoreness	-0.04	0.02	0.24	0.03	0.20	0.24	0.33	0.21	0.23	0.34	0.00	-0.52	0.25	0.01	0.57	0.95	-0.45	-0.46	-0.45	-0.60	0.84	0.34						
24.nTriads	-0.03	0.01	0.20	0.03	0.17	0.21	0.28	0.19	0.16	0.28	0.00	-0.38	0.31	0.05	0.65	0.93	-0.41	-0.41	-0.41	-0.53	0.80	0.16	0.91					
25.nSH	-0.03	0.01	0.21	0.02	0.18	0.21	0.25	0.19	0.16	0.27	0.00	-0.30	0.40	0.10	0.64	0.91	-0.39	-0.41	-0.41	-0.51	0.81	0.05	0.82	0.96				
26.nEgoConnectivity	-0.01	0.01	-0.13	0.12	-0.11	-0.12	-0.09	-0.10	-0.19	-0.16	0.00	-0.04	-0.18	-0.24	-0.07	-0.08	0.15	0.20	0.20	0.20	-0.20	0.10	0.01	0.05	-0.01			
27.elrsf	0.25	-0.48	0.04	0.74	-0.10	0.02	0.32	-0.08	-0.12	-0.14	0.00	0.09	-0.19	-0.21	-0.18	-0.17	0.20	0.26	0.26	0.24	-0.21	0.05	-0.12	-0.14	-0.17	0.12		
28.elrcf	0.05	-0.10	0.43	-0.03	0.41	0.38	0.22	0.37	0.09	0.29	0.00	0.01	0.53	0.30	0.40	0.33	-0.24	-0.29	-0.29	-0.29	0.40	-0.02	0.23	0.29	0.36	-0.11	-0.14	
29.elrv	0.02	-0.03	0.35	-0.09	0.39	0.31	0.13	0.34	0.08	0.27	0.00	0.01	0.41	0.25	0.34	0.24	-0.19	-0.23	-0.24	-0.23	0.30	-0.01	0.18	0.22	0.27	-0.06	-0.18	0.92

Table 5.7: The Local Correlation Matrix. Gray cells indicate $p < 0.05$.

The correlation matrices and descriptive analysis thus leads to some initial considerations that must be taken before a proper model is run. First, the correlations across many of the variables are very high (many larger than .9). This may be problematic and can lead to multi-collinearity concerns if I were the run the models on the current independent variables. Second, in order to properly test the hypotheses argued in the previous chapter, I will need to conduct a factor analysis to reduce the variables to the 5-Fundamental Structural Dimensions of the Graph.

Next, the descriptive statistics of the variables seem to indicate very tight interquartile ranges, but the outside quartiles are very sparse. This leads to an indication that there may be extreme outliers, which may cause issues if we conduct a factor analysis and later a regression analysis. Hence, proper additional cleansing of the data is necessary. Last, given the nature of the data and the problem at hand, as well as how the data was generated, I anticipate that the error terms in each of the empirical specifications will have high levels of correlation. In such an instance, it would be wise to run a Seemingly Unrelated Regression (SUR) in order to avoid the issues at hand. The standard estimation approach for this is Feasible Generalized Least Squares, and hence, endogeneity and heteroscdastisity concerns are fairly tempered given the level of consistently and efficiency of this estimator.

5.2 Factor Analysis

5.2.1 1-Ego-Network and 2-Ego-Network Factor Analysis

Given the high levels of correlation across the proposed independent variables, as well as the need to explain the proposed constructs with the measured variables, I conducted an exploratory factor analysis on the all variables except the dependent risk variables and the alpha and beta variables (which are control variables). The initial analysis of the 2-ego-network and the 1-ego-network data, respectively, indicated that the optimal number of factors to consider was 5, as we can observe from the scree plot. At first, I conducted two separate EFA's on the respective data sets. However, doing so led to some inconsistencies. While most variables loaded onto similar factors, some for the

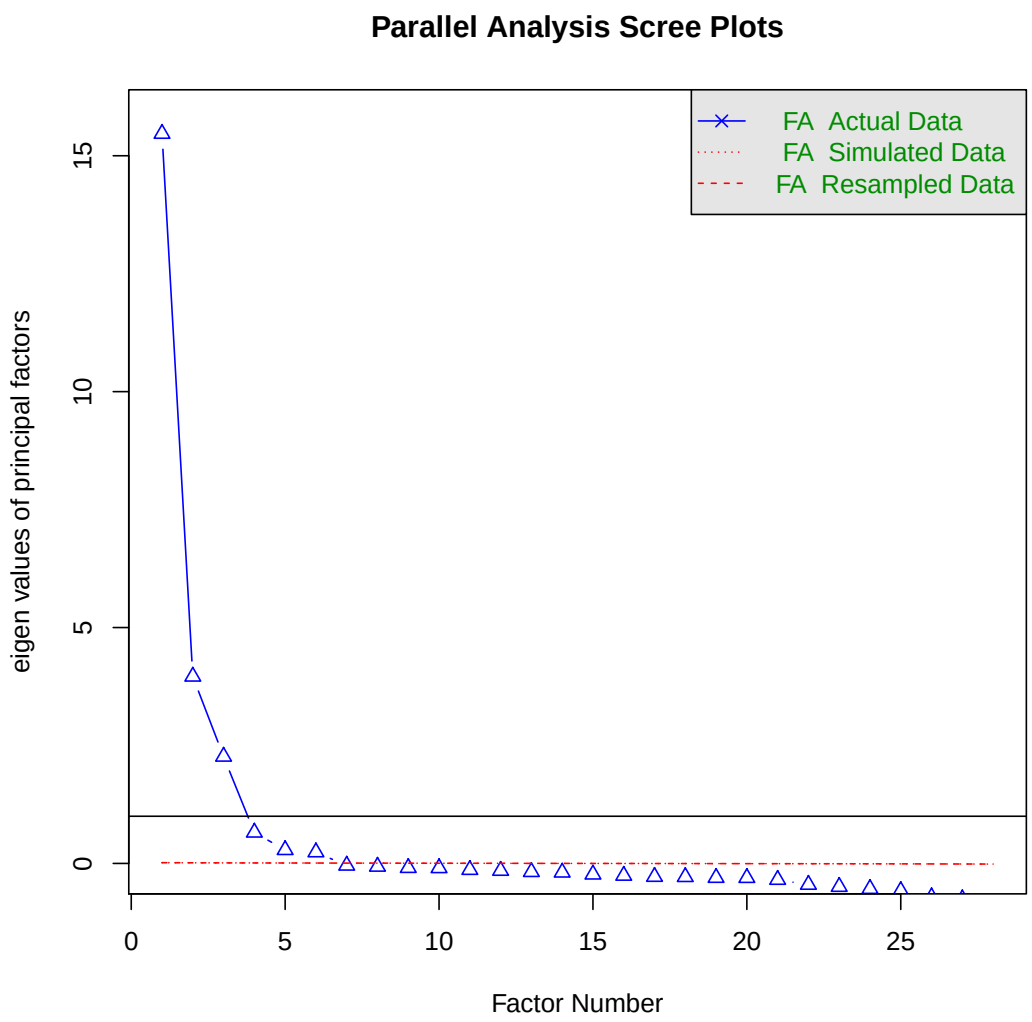


Figure 5.1: The scree plot for the EFA on the combined 2-Ego-Network and 1-Ego-Network data.

1-Ego-Network were loaded onto factors that did not seem to be clear as to why.

Given the nature of the data and the similarity and conformity of the 2-Ego-Network and 1-Ego-Network data, I had decided to pool the two data-sets together and run a joint-factor analysis to ensure a consistent loading of factors across the data sets. The scree plot for this final analysis is shown in Figure 5.1. Again, the optimal number of factors to consider was indicated to be 5. I had ran a final version of the EFA using the Minimum Residual Method with a Varimax Rotation in R using the *psych* package. The final loadings are shown in Table 5.8, with the highest loading of the

variable highlighted in gray. The recommended 1-to-1 mappings of the variables onto the factors is indicated in Figure 5.2.

	MR1	MR3	MR4	MR2	MR5
eNNodes	0.84	0.47	0.23	0.11	-0.01
eNArcs	0.82	0.27	0.49	0.12	0.01
eNSH	0.89	0.25	0.29	0.03	-0.01
eConstraintCen	0.84	0.37	0.32	0.09	-0.12
eAvgBetweeness	0.89	0.42	0.03	0.04	-0.02
eCenBetweeness	0.87	0.11	0.02	0.01	0.01
eCenEigen	0.88	0.44	0.12	0.06	-0.04
eCenCloseness	0.74	0.51	0.29	0.11	-0.07
eCenDegree	0.91	0.28	0.23	0.04	-0.03
eEccentricityCen	0.77	0.44	0.17	0.06	-0.06
eDiameter	0.39	0.86	0.11	0.07	-0.15
eRadius	0.42	0.75	0.10	0.06	-0.12
eNumOfLargestCliques	0.22	0.39	0.04	-0.25	0.10
eDensity	-0.29	-0.90	-0.14	-0.04	0.21
eAvgShortestPathDis	0.51	0.83	0.01	0.03	-0.12
eAvgCloseness	-0.06	-0.77	-0.28	-0.37	-0.02
eAvgEV	-0.41	-0.84	-0.16	-0.09	0.17
eAvgEccentricity	0.42	0.86	0.10	0.08	-0.16
eCliqueNum	0.35	0.15	0.76	0.47	0.04
eAvCliqueSize	0.30	0.11	0.85	0.38	0.10
eNTriads	0.65	0.04	0.67	0.08	0.02
eAvgConstraint	-0.19	-0.62	-0.68	-0.07	-0.02
eAvgDegree	0.27	0.26	0.81	0.40	0.18
eGirth	0.10	0.35	0.14	0.86	0.05
eClusterCoeff	0.04	-0.07	0.33	0.89	0.24
eClusterCoeffAv	0.09	0.03	0.32	0.85	0.29
eEdgeConnectivity	-0.09	-0.14	0.23	0.21	0.87
eNodeConnectivity	-0.03	-0.24	-0.03	0.21	0.72

Table 5.8: The 1-Ego-Network and 2-Ego-Network Factor Loadings from the EFA.

From the plot and the table, we can make the following observations. The variables Centralized Degree, Average Betweenness Centrality, Number of Structural Holes, Centralized Eigenvector Centrality, Centralized Betweenness Centrality, Centralized Constraint, Number of Arcs, Centralized Eccentricity, and Centralized Closeness Centrality have all loaded onto factor MR1. This factor mostly corresponds to the theoretical construct of Centrality, as I have argued earlier. What is surprising is that the number of structural holes have loaded onto this factor, which is a variable I had expected to load onto the clustering and possibly even feedback constructs. All variables loaded

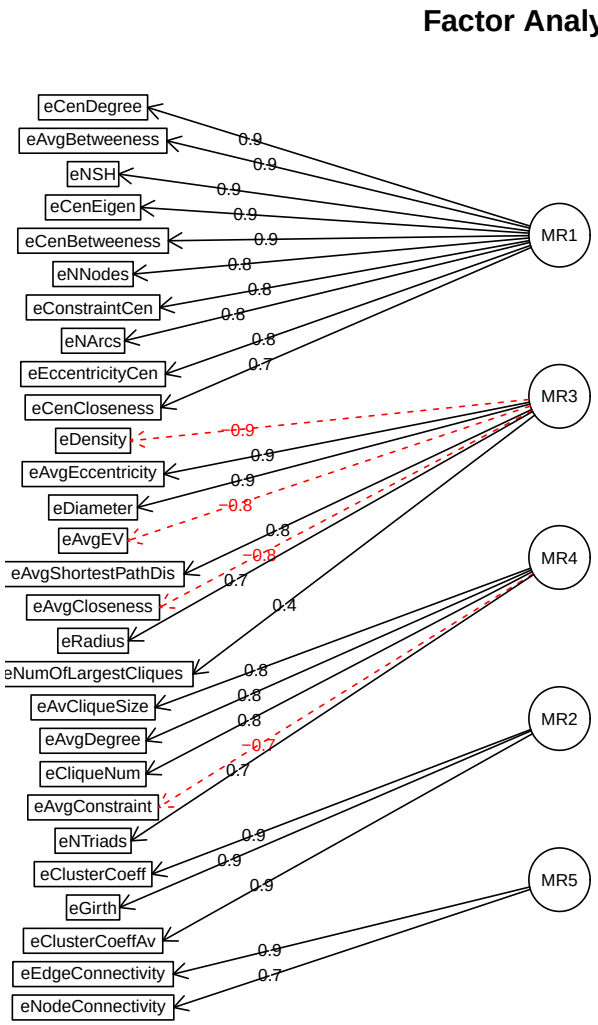


Figure 5.2: The final loadings of the 2-Ego-Network and 1-Ego-Network data.

positively onto this factor.

The second factor, MR3, had Density, Average Eccentricity, Diameter, Average Eigenvector Centrality, Average Shortest Path Distance, Average Closeness Centrality and Radius load onto it. Most of these were expected, and this factor would describe the Distance construct. Density, Average Eigenvector Centrality and Average Closeness Centrality loaded negatively on this construct. This makes sense. As density increases, the number of arcs increase, and hence this expands the number of possible paths between two nodes, which would imply that a new shortest path would be available between two firms. That is, the distance between two nodes shrink as the density increases, simply due to the fact that there are more available paths to link the nodes. In addition, networks with higher levels of average eigenvector centrality would indicate that many nodes are again highly connected and dense. By the preceding argument, we should have expected to see this. Last, as closeness centrality increases, again, nodes are more accessible to each other via short paths, and hence, the overall distance of the network shrink.

We notice that Average Clique Size, Average Degree, Clique Number, Average Constraint, and Number of Triads are loaded onto the third factor, MR4. This factor would describe the Clustering construct, and all variables, with the exception of degree, have loaded as expected. Constraint is negatively associated with the factor. This is quite surprising, since constraint is a measure of clustering in itself. Higher levels of constraint typically translate to an increased number of triads, which implies that the clustering of the network would be higher. This would indicate that the construct of clustering is much more multifaceted that simply characterizing it by triads.

Next, we notice that the Centralized Clustering Coefficient, Girth, and the Average Clustering Coefficient are associated with factor MR2, which would indicate that this factor is associated with the construct of Feedback. All variables loaded positively on the factor. Higher girth would indicate longer shortest cycles, which would of course translate to the network being more "thick". Higher levels of the Clustering coefficients, both centralized and average, would indeed indicate higher levels of Feedback. When the clustering coefficients are higher, there exist more triads. Eventually, as the proportion

of possible triads increase, so too does the establishment of a cycle (and, a larger length cycle). Hence, the loadings are indeed in line with the definition of the construct, and emphasize that "triadic analysis" is in itself descriptive of the different tangential construct of Feedback, rather than Clustering.

Last we notice that, as expected, node and edge connectivity loaded on the factor MR5, which of course would be indicative of the Connectivity construct. Node connectivity loaded less than that of edge connectivity.

Hence, we have the following factor interpretations:

1. MR1 = Centrality
2. MR3 = Distance
3. MR4 = Clustering
4. MR2 = Feedback
5. MR5 = Connectivity

5.2.2 Node and Edge Factor Analysis

Next I ran a factor analysis on the Node/Edge level data set to reduce the high levels of correlation and to extract the factors of which the theory seems to indicate it's existence. The node and edge data were pooled together and all variables except the alpha, beta, isNode, and risk measures were used in the factor analysis. A preliminary analysis produced the scree plot shown in Figure 5.3, which indicates that the optimal number of factors was again 5, inline with the theoretical justification.

We notice from Figure 5.4 as well as from Table 5.9 that there are five factors of which we must identify. The first factor loading, MR1, had the variables Number of Structural Holes, Number of Triads, Degree, Coreness, Density, Eigenvector Centrality, and Betweenness Centrality load onto it. Given that Degree, Coreness, Eigenvector and Betweenness Centrality loaded onto this factor indicates that this is our Centrality construct. What is interesting is the the number of structural holes and triads, typically Clustering and Feedback based measures, loaded onto this. However, given that this

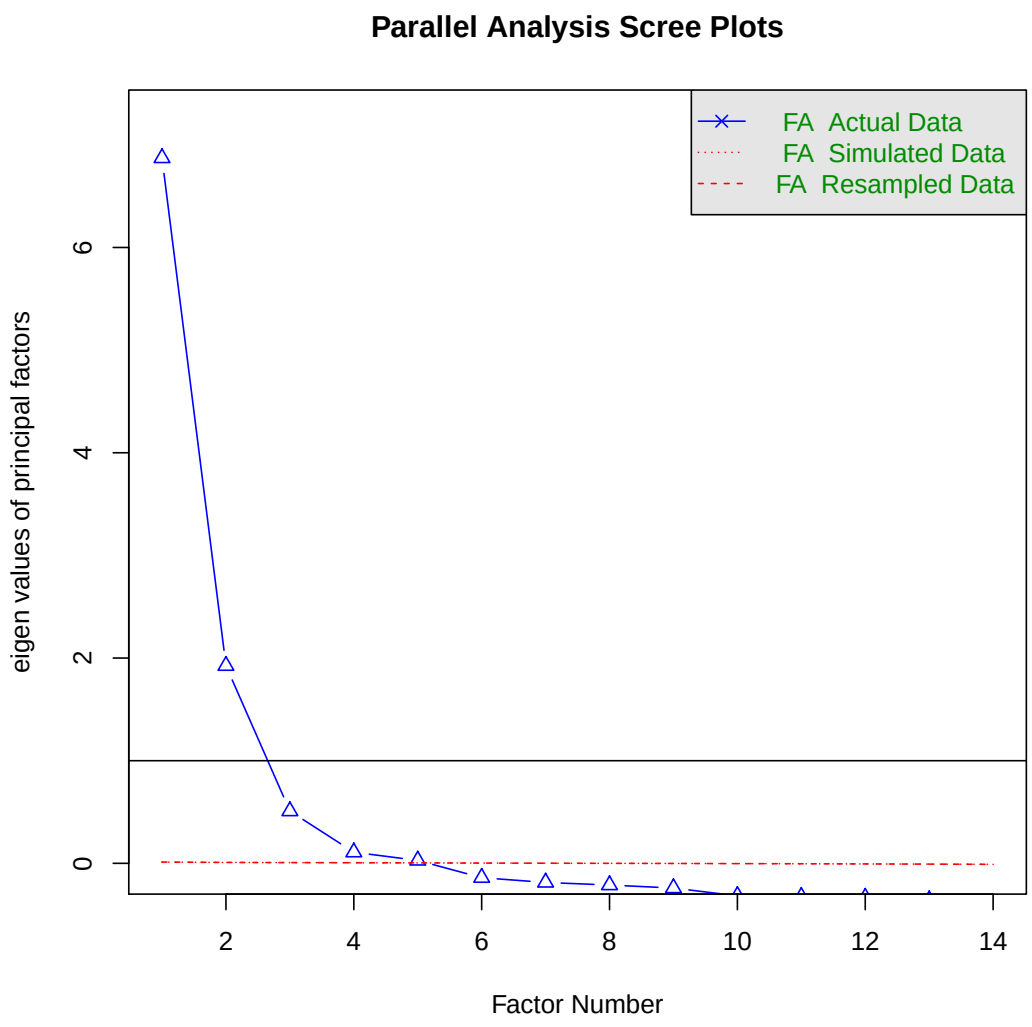


Figure 5.3: The scree plot for the EFA on the combined Node and Edge data.

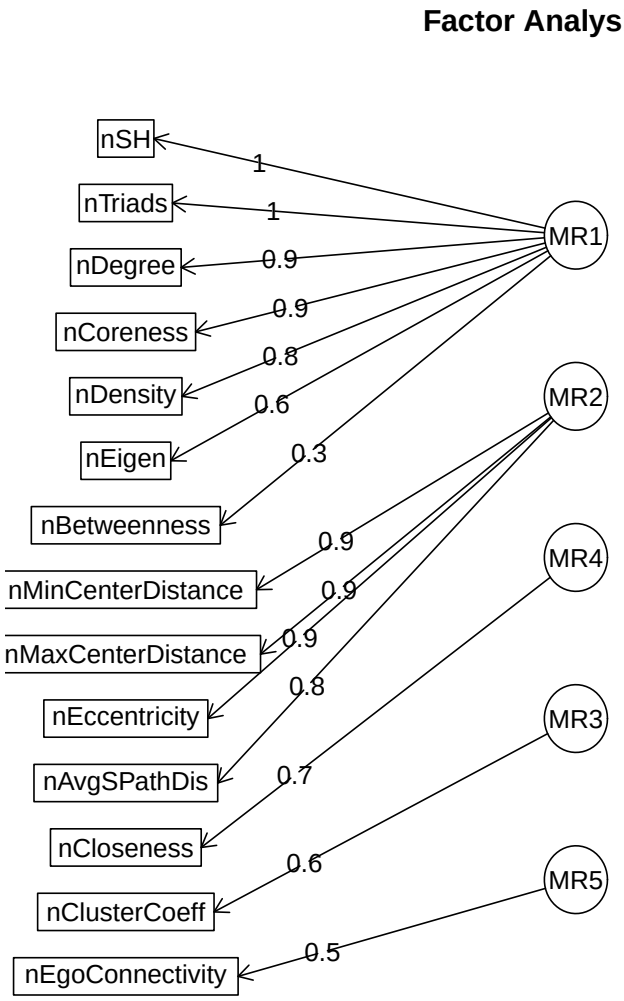


Figure 5.4: The final factor loadings for the Node and Edge data.

	MR1	MR2	MR4	MR3	MR5
nBetweenness	0.35	-0.20	0.10	-0.16	-0.30
nEigen	0.62	-0.27	0.37	-0.02	-0.05
nDegree	0.92	-0.26	-0.07	0.21	-0.17
nDensity	0.83	-0.24	0.23	0.23	-0.38
nCoreness	0.86	-0.25	-0.07	0.42	0.02
nTriads	0.96	-0.21	-0.00	0.06	0.15
nSH	0.96	-0.20	-0.01	-0.13	0.01
nEccentricity	-0.23	0.88	-0.09	-0.10	0.08
nMinCenterDistance	-0.25	0.91	-0.14	-0.05	0.16
nMaxCenterDistance	-0.25	0.91	-0.16	-0.04	0.17
nAvgSPathDis	-0.38	0.81	-0.15	-0.19	0.18
nCloseness	-0.01	-0.52	0.72	-0.15	-0.25
nClusterCoeff	0.10	-0.10	-0.05	0.56	0.11
nEgoConnectivity	0.01	0.14	-0.06	0.08	0.47

Table 5.9: The EFA Loadings on the Node/Edge data set

analysis is conducted at the local level, rather than a network-level, it can be argued that the number of possible triads and structural holes a node can have is inherently degree-based, which of course is centrality based. More generally this further illustrates the need for the network Constructs, since their observables are dependent on the structural level-of-analysis (That is, number of triads describes one factor at one structural level, but describes a different factor better at a different level).

The second factor, MR2, had the variables Minimum Center Distance, Maximum Center Distance, Eccentricity, and Average Shortest Path Distance load. This clearly is our Distance construct. All measures are indeed inherently distance-based.

The third factor MR4 was in a one-to-one relationship with the closeness centrality. I argue that this is the Clustering construct. When closeness has a low value, this indicates that nodes are less accessible, which implies that their ability to form cliques decreases. Therefore, at the node/edge level, it is reasonable to presume that this variable will load onto this factor.

The factor MR3 is our Feedback construct. Again, clustering coefficients are calculated as the proportion of actual triads to the number of possible triads that a node be part of. When this value is higher, then there are more triads, which increases the probability of a cycle forming, along with the cycle distance if one indeed forms. The

last factor, MR5, is obviously our Connectivity construct. Again, the k-ego connectivity indicates the number of ego networks one must "go up", or remove, before the entire network becomes disconnected. Hence, we can conclude the following:

1. MR1 = Centrality
2. MR2 = Distance
3. MR4 = Clustering
4. MR2 = Feedback
5. MR5 = Connectivity

After the factor analysis was conducted on the the primary independent variables of which needed to be reduced dimensionally, I used R's *psych* package's functionality to obtain the finalized factor scores. After, I conducted a confirmatory factor analysis (CFA) of the risk variables. Unlike the other independent variables, where the structural constructs can be argued to be described by the same observables, but at different levels of analysis, and there is reason to believe that a variable could be loaded on to more than one factor, the risk variables themselves should indeed be forced to load onto their own respective factors as the theory would mandate.

For each construct of risk (confidence, significance, and velocity), I had three operationalizations (average, standard deviation, and coefficient of variation). While I kept these for the primary analysis of the data, when they are used as the independent variable for the Ego-1 and Node/Edge models, they will exhibit high levels of correlation (as is indicative of the original correlation matrices). Therefore, I had ran a confirmatory factor analysis on the risk data for the 1-Ego-Network data (2-Ego-Network Risk) and the Local-Network data (1-Ego-Network Risk), respectively.

The statistics for the CFA for the 2-Ego-Network Risk are shown in Table 5.10. Unfortunately given the CFI, TLI, RMSEA, and SRMR, we notice this is not what traditionally would be considered a "good fit". However, I had made the decision to keep the loadings so as to reduce these observables into a single factor score for the

Measure	Value
Optimization method	NLMINB
Number of free parameters	21
Number of observations	254331
Estimator	ML
Model Fit Test Statistic	805754.964
Degrees of freedom	24
P-value (Chi-square)	0.000
Comparative Fit Index (CFI)	0.746
Tucker-Lewis Index (TLI)	0.619
Loglikelihood user model (H0)	-4278141.883
Loglikelihood unrestricted model (H1)	-3875264.401
Number of free parameters	21
Akaike (AIC)	8556325.766
Bayesian (BIC)	8556545.140
Sample-size adjusted Bayesian (BIC)	8556478.401
RMSEA	0.363
90 Percent Confidence Interval	0.363 , 0.364
P-value RMSEA $\leq$ 0.05	0.000
SRMR	0.254

Table 5.10: The 2-Ego-Network Risk CFA Statistics

independent variable in the model. While the reliability of these measures may be questioned, there is little doubt that the validity of them are in tact. The loads for the CFA are shown in Table 5.11.

I then ran a second CFA but on the 1-Ego-Network Risk Data (the independent variables for the Local-Network Risk Models). The results, shown in Table 5.12, were not too different from that of the 2-Ego-Network Risk Data. In fact, we notice that the statistics and loadings are very similar to each other, despite the data sets representing risk at different levels of analysis. With that said, we also notice the same issues with overall "fit". The primary issue at hand is of course reflected within the fit statistics of CFI, TLI, RMSEA, and SRMR. Again, I had chosen to ignore what would be considered a "poor fit" given the theoretical nature of the variables. The primary purpose at hand for the CFA is to (1) combine the known observables of each respective dependent variable into a single factor score and (2) reduce the potential of multi-collinearity in the final regression model. Despite the "poor fit", I argue that both of these goals have been accomplished, and hence, reliability should be of no concern for these specific

Variable	Loading
contribution	
avg_elrcf_2ego	1.000
sd_elrcf_2ego	2.188
cv_elrcf_2ego	0.034
significance	
avg_elrsf_2ego	1.000
sd_elrsf_2ego	-0.498
cv_elrsf_2ego	-9.351
velocity	
avg_elrv_2ego	1.000
sd_elrv_2ego	2.731
cv_elrv_2ego	-0.000

Table 5.11: The 2-Ego-Network Risk CFA Loadings

measures. The final loadings of this CFA are shown in Table 5.13

After, R's *lavaan* package was able to automatically compute the factor scores for these constructs. These scores, along with the factor scores from the EFA on the primary constructs, were organized into three separate data sets of which were then further cleansed of extreme outliers before the final analysis was conducted. The correlation tables for these final data sets are reported for the 2-Ego-Network data in Table 5.14, for the 1-Ego-Network data in Table 5.15, and last for the Local-Network data in Table 5.16. In addition, the summary statistics for the final data sets are reported in Table 5.17 (2-Ego-Network Data), Table 5.18 (1-Ego-Network Data), and Table 5.19 (Local-Network Data).

Measure	Value
Optimization method	NLMINB
Number of free parameters	21
Number of observations	322000
Estimator	ML
Model Fit Test Statistic	837850.243
Degrees of freedom	24
P-value (Chi-square)	0.000
Comparative Fit Index (CFI)	0.780
Tucker-Lewis Index (TLI)	0.670
Loglikelihood user model (H0)	-5548765.932
Loglikelihood unrestricted model (H1)	-5129840.811
Number of free parameters	21
Akaike (AIC)	11097573.865
Bayesian (BIC)	11097798.193
Sample-size adjusted Bayesian (BIC)	11097731.454
RMSEA	0.329
90 Percent Confidence Interval	0.329,0.330
P-value RMSEA $\leq$ 0.05	0.000
SRMR	0.167

Table 5.12: The 1-Ego-Network Risk CFA Statistics

Variable	Loading
contribution	
avg_elrcf_ego	1.000
sd_elrcf_ego	1.831
cv_elrcf_ego	0.025
significance	
avg_elrsf_ego	1.000
sd_elrsf_ego	-0.428
cv_elrsf_ego	-0.214
velocity	
avg_elrv_ego	1.000
sd_elrv_ego	2.111
cv_elrv_ego	-0.000

Table 5.13: The 1-Ego-Network Risk CFA Loadings

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1.alpha																				
2.beta	0.07																			
3.Centrality	-0.08	0.15																		
4.Distance	0.00	0.03	-0.09																	
5.Clustering	-0.09	0.10	0.16	-0.29																
6.Feedback	-0.07	0.12	-0.06	-0.21	0.02															
7.Connectivity	0.02	-0.01	0.16	0.47	-0.30	-0.53														
8.avg_elrcf	0.14	-0.21	0.40	0.06	0.13	0.01	0.07													
9.avg_elrsf	0.31	-0.55	-0.18	-0.01	0.02	0.13	-0.09	0.01												
10.avg_elrv	-0.03	0.11	0.50	0.06	0.12	-0.04	0.11	0.82	-0.41											
11.sd_elrcf	0.10	-0.14	0.49	0.08	0.13	-0.01	0.10	0.94	-0.09	0.86										
12.sd_elrsf	0.21	-0.40	0.36	0.07	0.28	0.17	-0.01	0.55	0.53	0.25	0.52									
13.sd_elrv	-0.01	0.07	0.56	0.07	0.12	-0.05	0.13	0.79	-0.37	0.96	0.89	0.32								
14.cv_elrcf	-0.05	0.10	0.44	0.15	0.07	-0.04	0.14	0.38	-0.30	0.50	0.62	0.19	0.63							
15.cv_elrsf	-0.11	0.17	0.62	0.07	0.27	0.06	0.10	0.58	-0.43	0.78	0.66	0.45	0.79	0.54						
16.cv_elrv	-0.01	0.01	0.12	0.11	0.02	-0.03	0.07	0.13	-0.14	0.10	0.16	0.10	0.12	0.27	0.19					
17.Centrality2	-0.07	0.11	0.85	-0.17	0.20	-0.06	0.06	0.28	-0.12	0.36	0.35	0.29	0.41	0.29	0.47	0.05				
18.Distance2	0.00	0.02	-0.13	0.87	-0.34	-0.13	0.44	0.01	0.01	0.00	0.03	0.03	0.02	0.10	0.01	0.09	-0.17			
19.Clustering2	-0.08	0.11	0.37	-0.05	0.39	0.02	0.02	0.19	-0.07	0.23	0.23	0.24	0.25	0.21	0.34	0.06	0.25	-0.02		
20.Feedback2	0.06	-0.12	-0.25	-0.15	-0.05	-0.31	0.01	-0.15	0.04	-0.16	-0.16	-0.18	-0.17	-0.15	-0.25	-0.07	-0.12	-0.14	-0.15	
21.Connectivity2	-0.02	0.01	0.02	-0.38	0.14	0.39	-0.53	0.03	0.02	0.01	0.01	0.03	0.00	-0.03	0.01	-0.02	0.02	-0.31	0.13	0.12

Table 5.14: The 2-Ego-Network Factored Correlation Matrix. Gray cells indicate $p < 0.05$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1.alpha																							
2.beta	0.08																						
3.contribution	0.18	-0.36																					
4.significance	0.09	-0.10	-0.12																				
5.velocity	0.03	-0.04	0.42	0.10																			
6.Centrality	0.04	-0.07	-0.05	0.10	-0.02																		
7.Distance	-0.02	0.06	0.07	-0.08	0.15	-0.59																	
8.Clustering	-0.09	0.11	0.10	-0.19	0.14	-0.05	0.07																
9.Feedback	-0.07	0.10	0.02	-0.12	-0.05	-0.31	-0.08	-0.12															
10.Connectivity	-0.03	0.05	0.04	-0.07	0.04	0.03	0.14	0.16	-0.15														
11.avg_elrcf	0.13	-0.21	0.42	-0.14	0.38	-0.08	0.24	0.26	0.01	0.09													
12.avg_elrsf	0.29	-0.53	0.23	0.10	-0.28	-0.04	-0.08	-0.05	0.12	0.00	-0.01												
13.avg_elrv	-0.02	0.07	0.20	-0.29	0.44	-0.02	0.23	0.24	-0.04	0.07	0.79	-0.40											
14.sd_elrcf	0.11	-0.16	0.44	-0.18	0.43	-0.05	0.25	0.28	-0.02	0.10	0.93	-0.09	0.82										
15.sd_elrsf	0.21	-0.44	0.42	-0.14	0.15	-0.16	0.16	0.41	0.17	0.15	0.45	0.50	0.14	0.42									
16.sd_elrv	-0.01	0.06	0.23	-0.30	0.47	0.00	0.24	0.25	-0.06	0.08	0.74	-0.38	0.96	0.84	0.17								
17.cv_elrcf	-0.02	0.06	0.25	-0.23	0.29	-0.02	0.20	0.22	-0.05	0.09	0.30	-0.23	0.39	0.55	0.14	0.52							
18.cv_elrsf	-0.07	0.09	0.19	-0.44	0.44	-0.11	0.26	0.47	0.07	0.15	0.50	-0.43	0.67	0.56	0.47	0.68	0.41						
19.cv_elrv	-0.02	0.02	0.11	-0.10	0.18	-0.05	0.13	0.12	-0.01	0.04	0.20	-0.19	0.18	0.23	0.10	0.19	0.25	0.27					
20.Centrality2	-0.04	0.07	0.05	-0.10	0.04	-0.93	0.59	0.07	0.23	0.12	0.11	0.03	0.06	0.08	0.17	0.04	0.05	0.15	0.06				
21.Distance2	0.02	-0.04	-0.03	0.05	-0.07	0.50	-0.57	-0.03	-0.09	0.07	-0.12	0.03	-0.10	-0.11	-0.07	-0.10	-0.07	-0.11	-0.06	-0.43			
22.Clustering2	-0.09	0.10	0.04	-0.16	0.07	0.07	-0.08	0.79	-0.03	0.16	0.15	-0.02	0.14	0.16	0.31	0.15	0.12	0.34	0.06	-0.05	0.02		
23.Feedback2	0.04	-0.08	-0.04	0.09	-0.11	0.24	-0.45	-0.17	-0.31	-0.08	-0.19	0.07	-0.18	-0.18	-0.17	-0.18	-0.14	-0.26	-0.10	-0.18	0.36	-0.18	
24.Connectivity2	-0.03	0.04	0.03	-0.05	0.02	-0.04	0.03	0.14	0.12	0.63	0.07	0.02	0.05	0.07	0.12	0.05	0.06	0.10	0.02	0.12	0.00	0.11	-0.17

Table 5.15: The 1-Ego-Network Factored Correlation Matrix. Gray cells indicate $p < 0.05$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1.alpha																		
2.beta	0.05																	
3.isNode	-0.02	0.05																
4.contribution	0.13	-0.27	-0.19															
5.significance	0.05	-0.05	0.24	-0.25														
6.velocity	0.03	-0.07	-0.17	0.62	-0.16													
7.Centrality	-0.02	0.00	-0.31	0.10	-0.14	0.07												
8.Distance	-0.01	0.02	0.06	-0.13	0.11	-0.10	0.10											
9.Clustering	0.06	-0.02	0.51	-0.11	0.14	-0.06	-0.30	0.02										
10.Feedback	-0.01	0.00	-0.56	0.14	-0.22	0.10	0.36	-0.07	-0.09									
11.Connectivity	-0.02	0.04	0.03	-0.07	0.12	-0.05	-0.43	0.01	0.02	-0.04								
12.elrsf	0.27	-0.49	-0.01	0.21	0.05	-0.01	-0.01	0.16	-0.12	0.06	0.05							
13.elrcf	0.14	-0.27	0.04	0.20	-0.05	0.10	0.32	-0.10	0.03	0.01	-0.51	0.15						
14.elrv	-0.01	0.01	0.00	0.09	-0.12	0.11	0.33	-0.18	0.06	0.01	-0.51	-0.28	0.75					
15.Centrality2	-0.01	0.01	-0.12	0.05	-0.04	0.03	0.61	-0.18	-0.09	-0.03	0.09	-0.07	0.13	0.16				
16.Distance2	-0.01	0.01	0.02	0.00	0.02	0.00	0.11	0.37	-0.02	-0.04	0.02	0.06	-0.04	-0.05	-0.06			
17.Clustering2	-0.06	0.01	-0.21	0.14	-0.18	0.07	0.39	-0.11	-0.35	0.28	-0.26	0.07	0.19	0.16	0.08	-0.01		
18.Feedback2	0.01	-0.01	0.00	0.00	0.03	0.00	0.14	-0.04	0.14	0.37	0.10	-0.03	-0.08	-0.06	0.00	-0.02	0.00	
19.Connectivity2	0.01	0.01	0.15	-0.01	0.03	0.01	0.13	0.04	0.24	-0.15	-0.35	-0.07	0.29	0.31	0.03	-0.01	0.00	-0.15

Table 5.16: The Local-Network Factored Correlation Matrix. Gray cells indicate $p < 0.05$.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
alpha	90,442	3.077	1.152	1	2	4	7
beta	90,442	5.741	2.743	1	4	8	10
Centrality	90,442	0.135	1.103	−1.010	−0.550	0.421	4.617
Distance	90,442	0.881	0.734	−0.923	0.358	1.507	2.020
Clustering	90,442	−0.137	0.706	−1.710	−0.562	0.123	2.517
Feedback	90,442	0.117	0.712	−1.864	−0.230	0.594	1.419
Connectivity	90,442	−0.210	0.470	−1.255	−0.565	0.138	1.022
avg_elrcf	90,442	8.883	6.321	1.105	3.998	12.312	40.029
avg_elrsf	90,442	5.707	2.808	1.593	3.601	7.205	16.264
avg_elrv	90,442	3.357	4.550	−0.787	0.131	5.206	21.328
sd_elrcf	90,442	15.488	12.606	1.237	5.359	23.036	59.365
sd_elrsf	90,442	1.595	0.785	0.462	0.994	2.029	4.205
sd_elrv	90,442	10.199	11.481	0.342	1.765	15.110	45.107
cv_elrcf	90,442	1.618	0.460	0.784	1.249	1.945	2.836
cv_elrsf	90,442	0.309	0.140	0.104	0.197	0.397	0.688
cv_elrv	90,442	2.405	4.728	−18.453	1.803	3.950	24.609

Table 5.17: Summary Statistics for The Factored 2-Ego-Network Data-Set.

A final data table was then assembled and cleaned for further preprocessing. The variables that were intended to be reduced dimensionally were removed and replaced with the construct factor scores. To ensure the data was proper and free of outliers for analysis, I removed the outer 5% of the data (2.5% on each side). Unfortunately the factor loadings and the original simulation had produced some very large outliers that could not be handled during the analysis phase in any other manner. Hence, they were removed.

5.3 Structural Difference Testing

As I mentioned in the hypothesis development, we can and should expect the 2-Ego-Networks and the 1-Ego-Networks to have different structures along the five structural dimensions. Figure 5.5 illustrates these differences by each construct for each level of analysis (2-Ego, 1-Ego, and Local).

We notice from these plots that the structural differences are indeed there, and the structure itself appears to depend on the specific level of analysis. We also notice that the speculation of Local structures being closer to that of the 2-Ego-Network

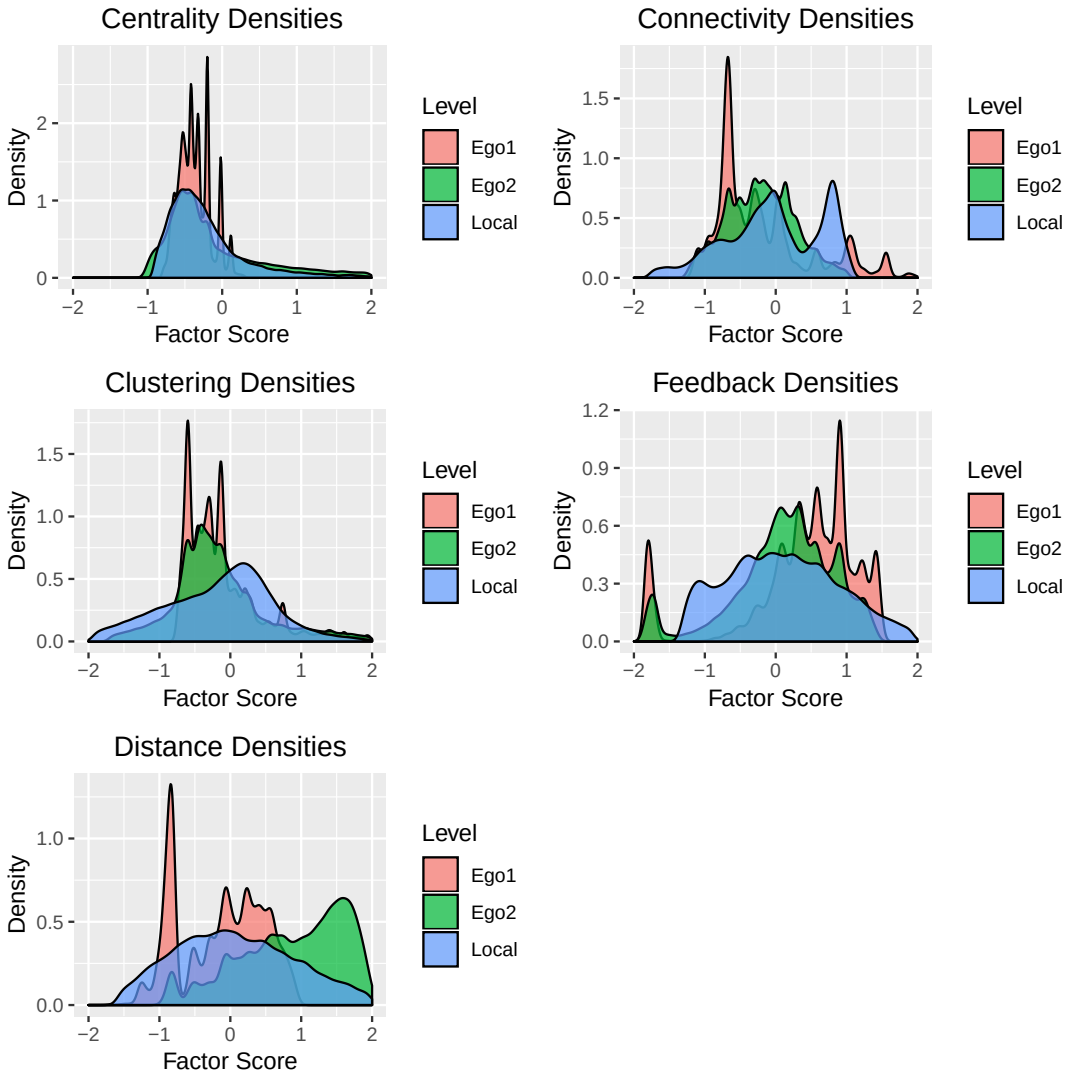


Figure 5.5: Densities for each Construct by Level of Analysis

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
alpha	101,035	3.090	1.166	1	2	4	7
beta	101,035	5.616	2.753	1	3	8	10
contribution	101,035	−0.543	11.118	−39.659	−7.587	2.877	216.320
significance	101,035	0.097	2.900	−33.277	−1.198	1.850	16.741
velocity	101,035	−0.404	6.814	−34.872	−4.538	2.072	45.083
Centrality	101,035	−0.360	0.211	−0.818	−0.526	−0.200	0.313
Distance	101,035	−0.138	0.589	−1.715	−0.822	0.350	0.942
Clustering	101,035	−0.032	0.712	−0.775	−0.491	0.133	3.480
Feedback	101,035	0.410	0.804	−1.805	0.128	0.906	1.485
Connectivity	101,035	−0.139	0.783	−1.140	−0.677	0.083	3.192
avg_elrcf	101,035	7.052	6.386	0.867	2.724	9.104	49.182
avg_elrsf	101,035	6.240	3.123	1.511	3.908	7.865	17.797
avg_elrv	101,035	1.612	3.546	−0.860	−0.474	2.046	26.660
sd_elrcf	101,035	9.539	9.755	0.744	2.761	12.511	54.381
sd_elrsf	101,035	1.288	0.672	0.387	0.781	1.629	3.711
sd_elrv	101,035	4.606	6.861	0.184	0.686	5.149	39.536
cv_elrcf	101,035	1.244	0.400	0.558	0.939	1.489	2.453
cv_elrsf	101,035	0.231	0.116	0.077	0.141	0.295	0.614
cv_elrv	101,035	1.063	4.466	−18.669	−0.996	2.975	21.160

Table 5.18: Summary Statistics for The Factored 1-Ego-Network Data-Set.

than to that of the 1-Ego-Network is reflected here, hence illustrating the argument posed earlier that "Local" structure is actually closer to "Global" structure, since local structure figures are calculated relative to the entire global network.

To test these hypotheses, I conducted 5 one-tailed difference of means tests to compare the constructs of the 2-Ego-Network structures against the 1-Ego-Network structures. In addition, I conducted a difference of means test between the 1-Ego-Network risk and the 2-Ego-Network by conducting a one-tailed hypothesis test. The same analysis was conducted for comparisons to the local. Last, I compared the structural characteristics between the edge's and the node's. These results are summarized in Table 5.20.

5.4 Causal Analysis

After the factor analysis as well as the data cleaning, I had 3 final cleaned data sets ready for analysis. The final correlation matrices and summary statistics, respectively,

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
alpha	205,327	3.067	1.175	1	2	4	7
beta	205,327	5.680	2.785	1	3	8	10
isNode	205,327	0.337	0.473	0	0	1	1
contribution	205,327	−1.261	12.097	−52.568	−6.823	−1.431	242.254
significance	205,327	0.119	2.501	−24.919	−0.576	1.561	23.822
velocity	205,327	−0.772	8.015	−24.950	−3.709	−2.304	88.491
Centrality	205,327	−0.193	0.583	−0.924	−0.567	−0.026	2.908
Distance	205,327	0.090	0.844	−1.593	−0.557	0.689	2.170
Clustering	205,327	−0.151	0.758	−1.945	−0.678	0.354	2.199
Feedback	205,327	0.099	0.775	−1.332	−0.494	0.657	1.948
Connectivity	205,327	−0.037	0.676	−1.790	−0.485	0.608	1.087
elrsf	205,327	6.549	3.494	1.322	3.958	8.335	18.874
elrcf	205,327	2.520	3.087	0.202	0.703	3.009	23.032
elrv	205,327	−0.480	0.767	−1.000	−0.923	−0.398	3.641

Table 5.19: Summary Statistics for The Factored Node/Edge Data-Set.

Construct	$Ego_1 < Ego_2$	$Local < Ego_1$	$Node < Edge$
Centrality	$p < 2.2e-16*$	$p = 1$	$p < 2.2e-16*$
Distance	$p < 2.2e-16*$	$p < 2.2e-16*$	$p = 1$
Clustering	$p = 1$	$p < 2.2e-16*$	$p = 1$
Feedback	$p = 1$	$p < 2.2e-16*$	$p < 2.2e-16*$
Connectivity	$p = 1$	$p = 1$	$p = 1$
Velocity	$p < 2.2e-16*$	$p < 2.2e-16*$	$p < 2.2e-16*$

Table 5.20: Results of One-Tailed Structural Comparison.

are reported for the 2-Ego-Network Data (Figures 5.14,5.17),1-Ego-Network Data (Figures 5.15,5.18),Local-Network Data (Figures 5.16,5.19). The densities of the different levels of analysis for all constructs at that level, the dependent risk variables, and the independent factored risk variables, are shown for the 2-Ego-Network(Figure 5.6), the 1-Ego-Network(Figures 5.7,5.8), and the Local-Network (Figures 5.8,5.9).

Before conducting the final analysis, we can also take a look at the raw data plots to gain some insight into the hypothesized relationships between the structural constructs and the risk. The plots shown in Figure 5.10 indicate that most of the associations appear to be curvilinear, with Connectivity and Distance appearing with a positive curvilinear association and with the other factors as a negative curvilinear association. In Figure 5.11, we notice that Centrality, Distance, and Connectivity appear to follow a positive curvilinear association, while Clustering and Feedback are negative curvilinear.

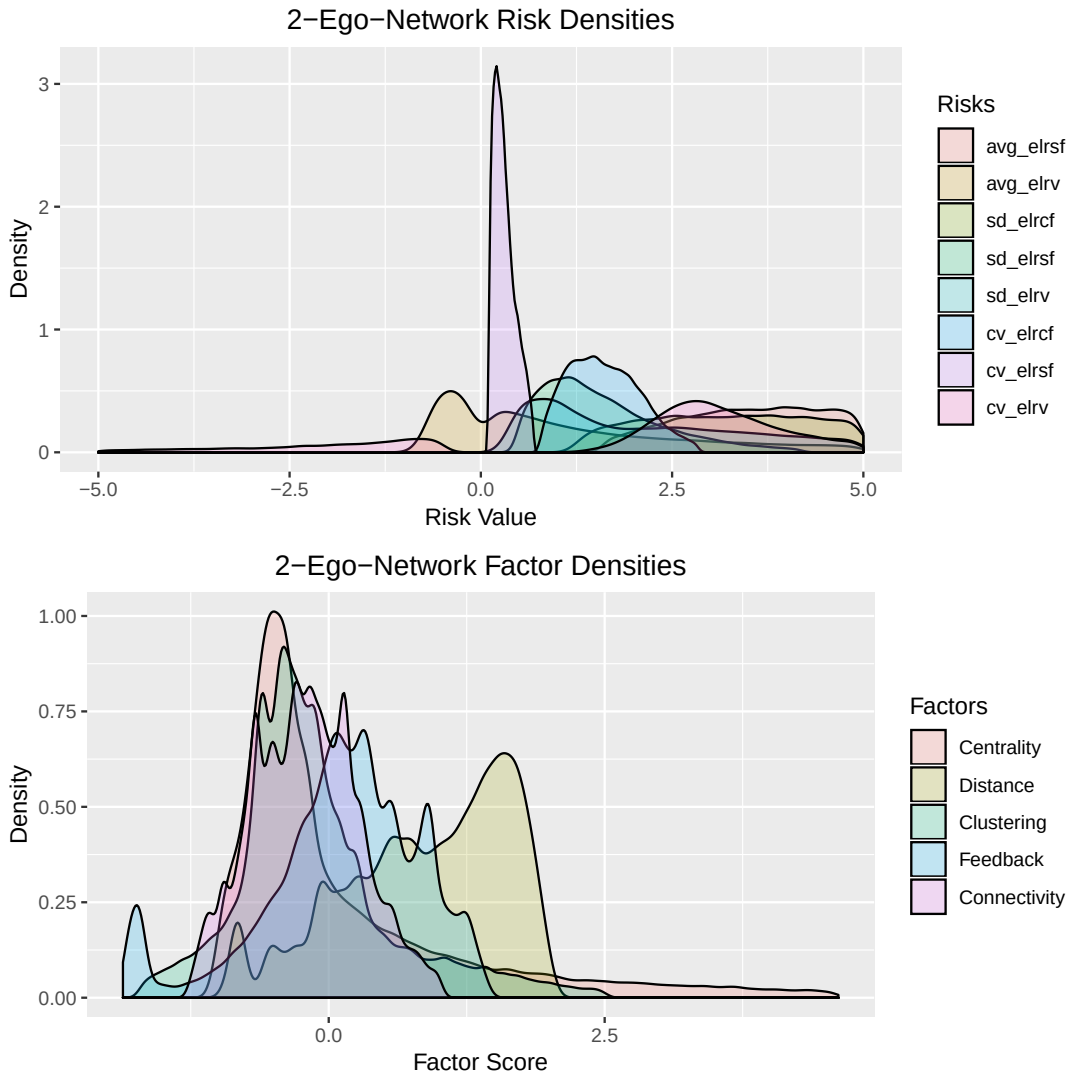


Figure 5.6: The 2-Ego-Network Densities for the dependent variables (above) and the constructs (below).

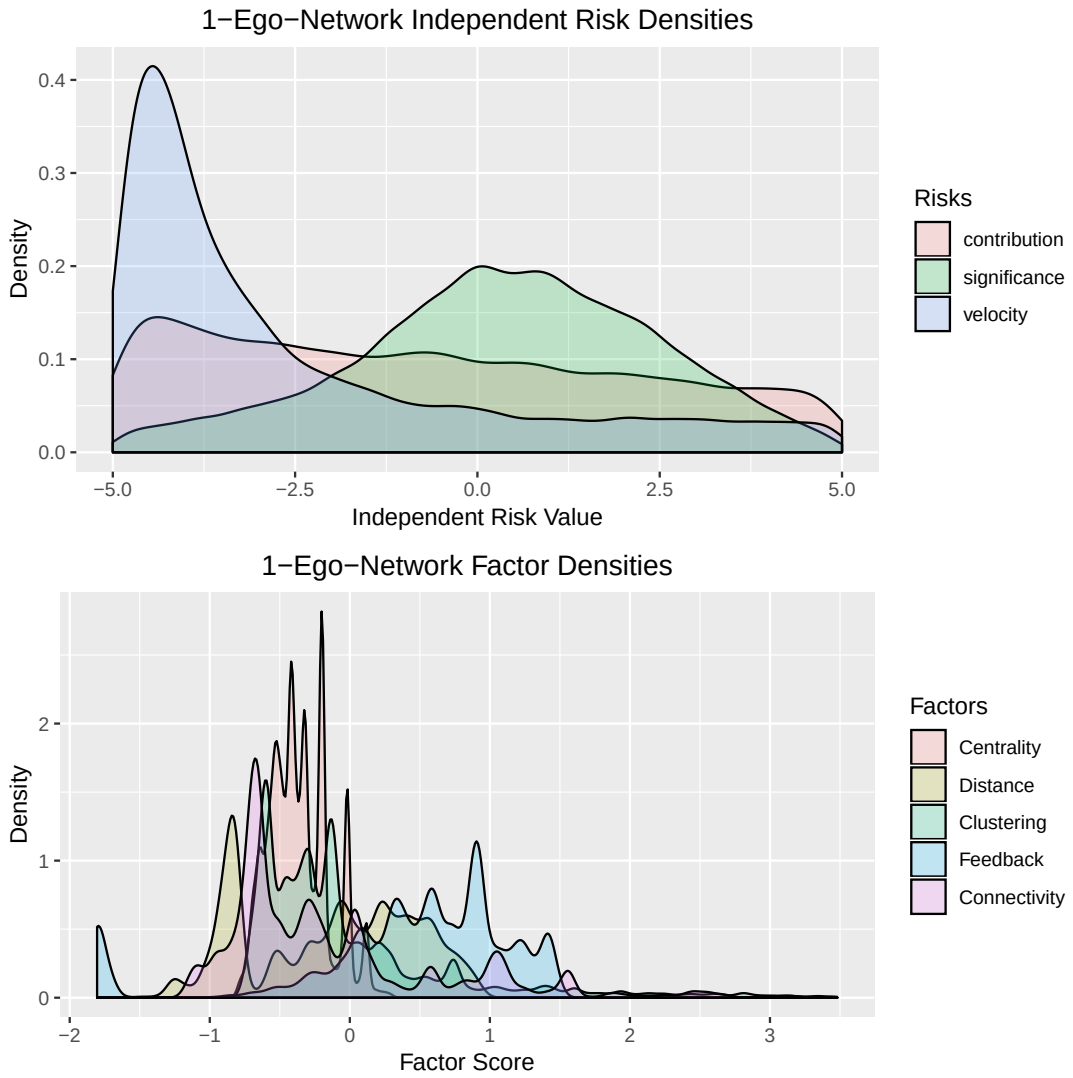


Figure 5.7: The 1-Ego-Network Densities for the independent variable risk factors (above) and the constructs (below).

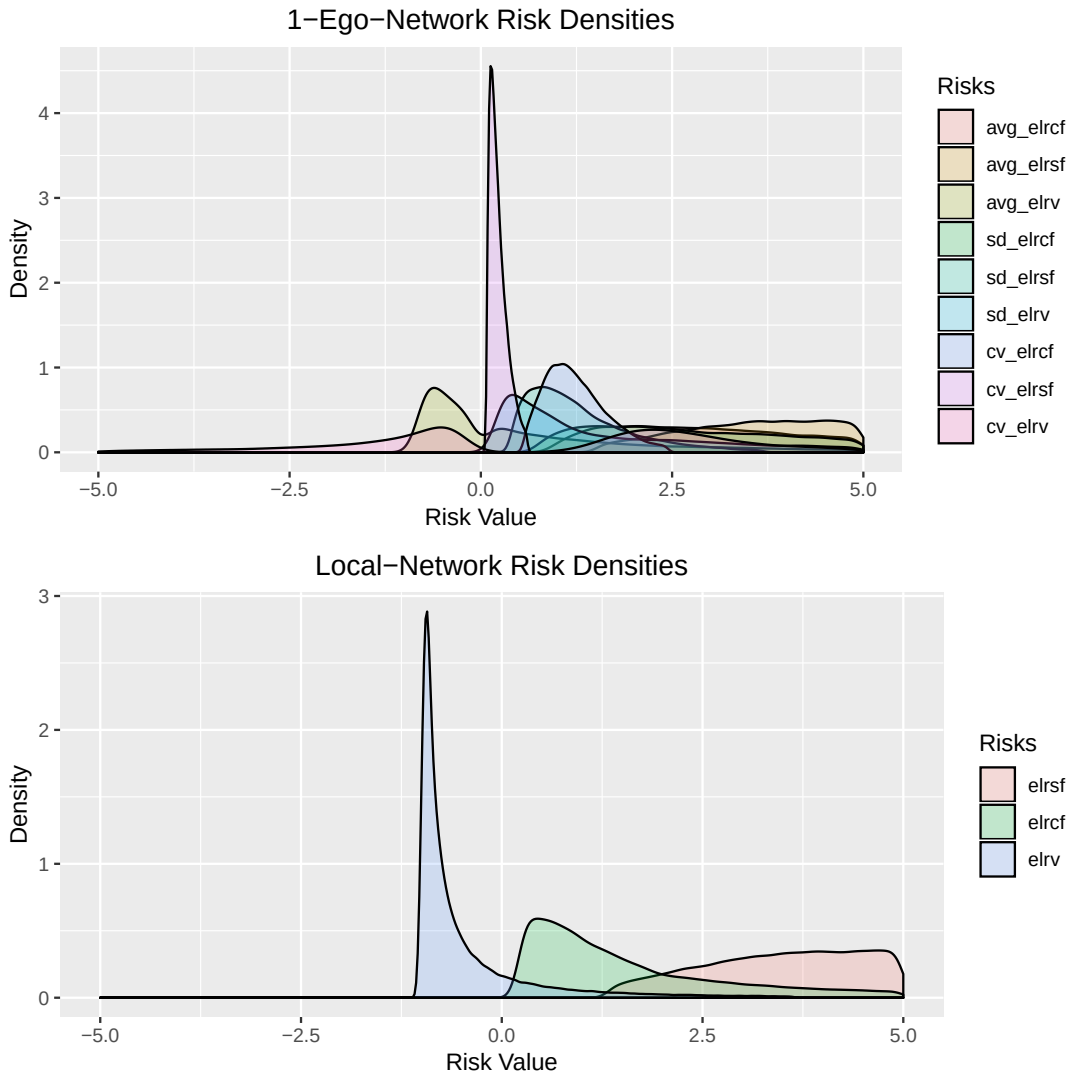


Figure 5.8: The 1-Ego-Network Densities for the dependent risk variables (above) and the Local dependent risk variables (below)

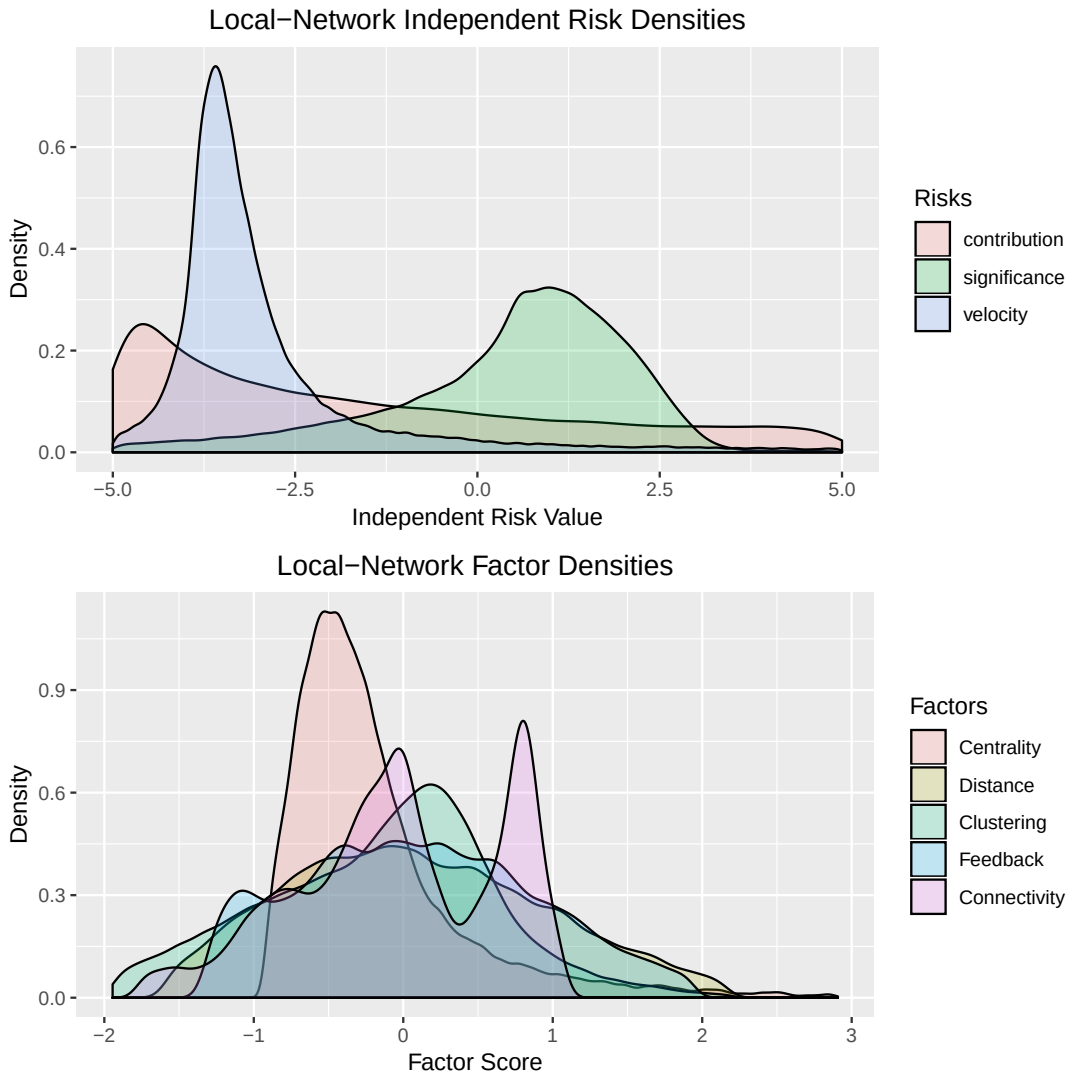


Figure 5.9: The Local Densities for the independent risk variables (above) and the local densities for the constructs

Last, in the local plots shown in Figure 5.13, we notice Centrality, and Feedback appear to follow a negative curvilinear association, while Clustering, Distance, and Connectivity appear to be positive curvilinear. Therefore, from a visual inspection, we can see that the results of our analysis below may justify most of the hypotheses above.

In regards to associations between higher-level risk and lower-level risk, we can further inspect visually the potential associations between these two. Figure 5.12 illustrates the potential association between the three risk constructs of the 2-Ego-Network and the risk of the 1-Ego-Network. We notice that contribution and significance are in negative curvilinear associations with each other, while velocity is in a positive association with risk. In Figure 5.14 we see the association between the local risk and the higher level 1-Ego-Network risk constructs. Contribution and significance appears to be in a negative association, while the velocity is only slightly in negative curvilinear association. I have hypothesized that these would be linear and have modeled them as such.

With this in mind, as well as the empirical models listed earlier, I ran three separate Seemingly Unrelated Regression models using R *systemfit* package. Before doing so, I had levered R's *usdm* package to conduct a step-wise VIF on all the independent variables for the three datasets. The first is shown in Table 5.21. We can see from the table that all VIFs for the planned regressors are less than 10, indicating that multicollinearity would not be of a concern in the final fitting of the model.

Next, I ran the same analysis on the 1-Ego-Network data set to isolate the potential issues of multicollinearity, with results shown in Table 5.22. We can see indeed that despite the poor fit on the CFA on the risk variables of contribution, significance and velocity, their respective VIF scores were very low, indicating that these indeed would not pose problems with multicollinearity. All other regressors had a VIF less than 10, which indicates again the multicollinearity should not be a cause of concern.

Last, I ran the same procedure on the local data set to once again ensure that there will be no issues with multicollinearity. As we can see from Table 5.23, once again, that despite the poor fit from the CFA on the risk variables of contribution, significance and velocity, that the VIFs were all low, indicating that these would not be a cause of

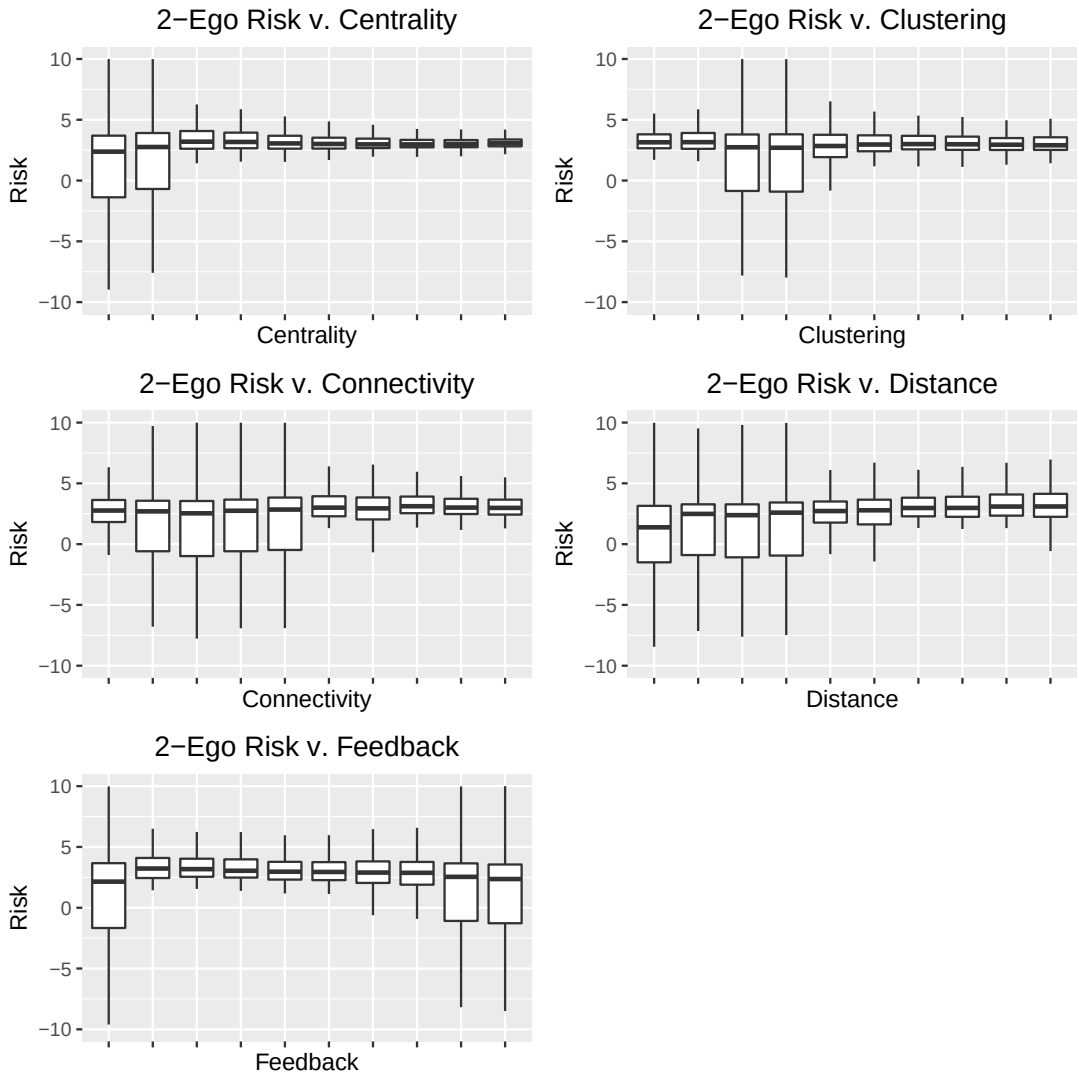


Figure 5.10: 2-Ego-Network Risk (CV of Velocity) plotted against 2-Ego-Network Structure Constructs

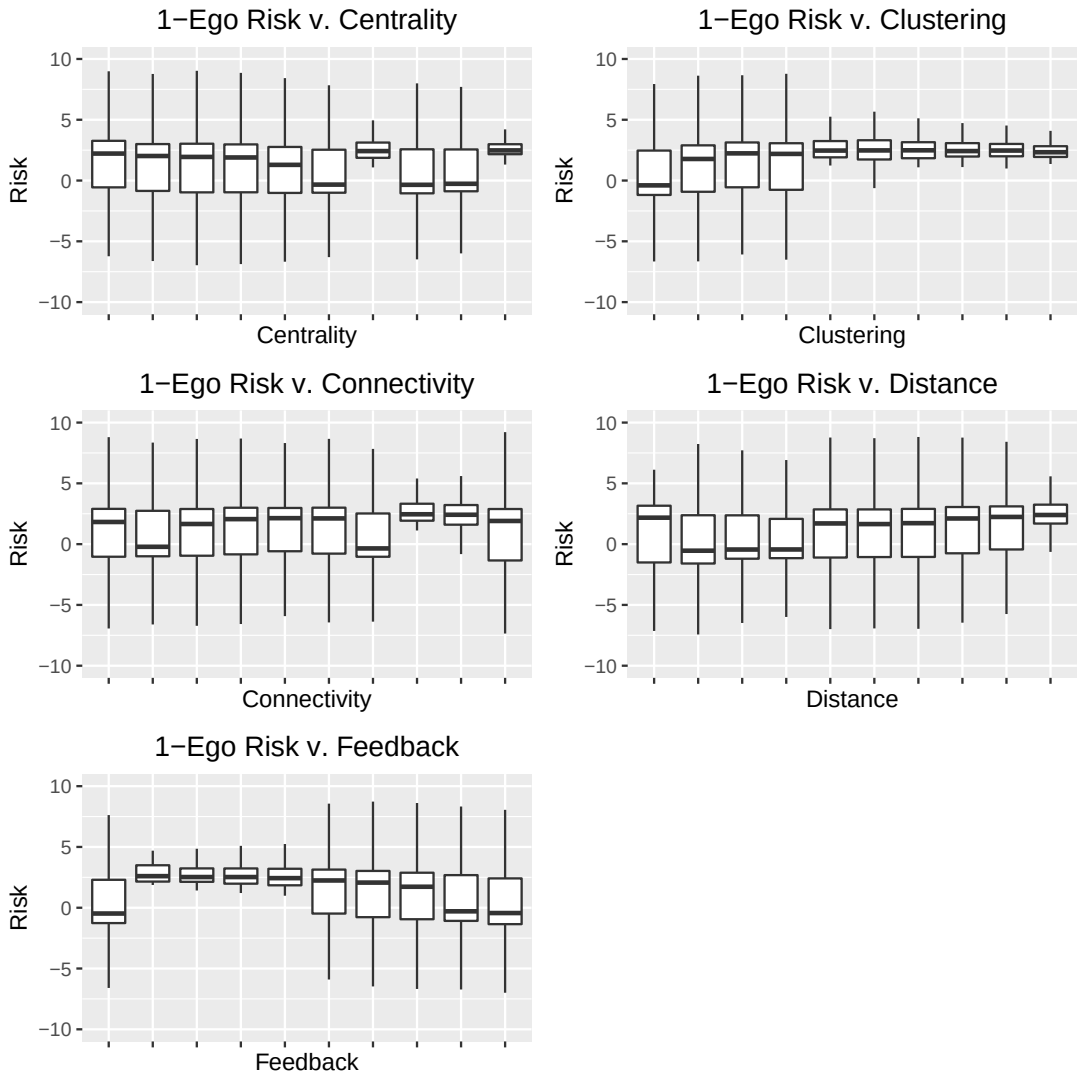


Figure 5.11: 1-Ego-Network Risk (CV of Velocity) plotted against 1-Ego-Network Structure Constructs

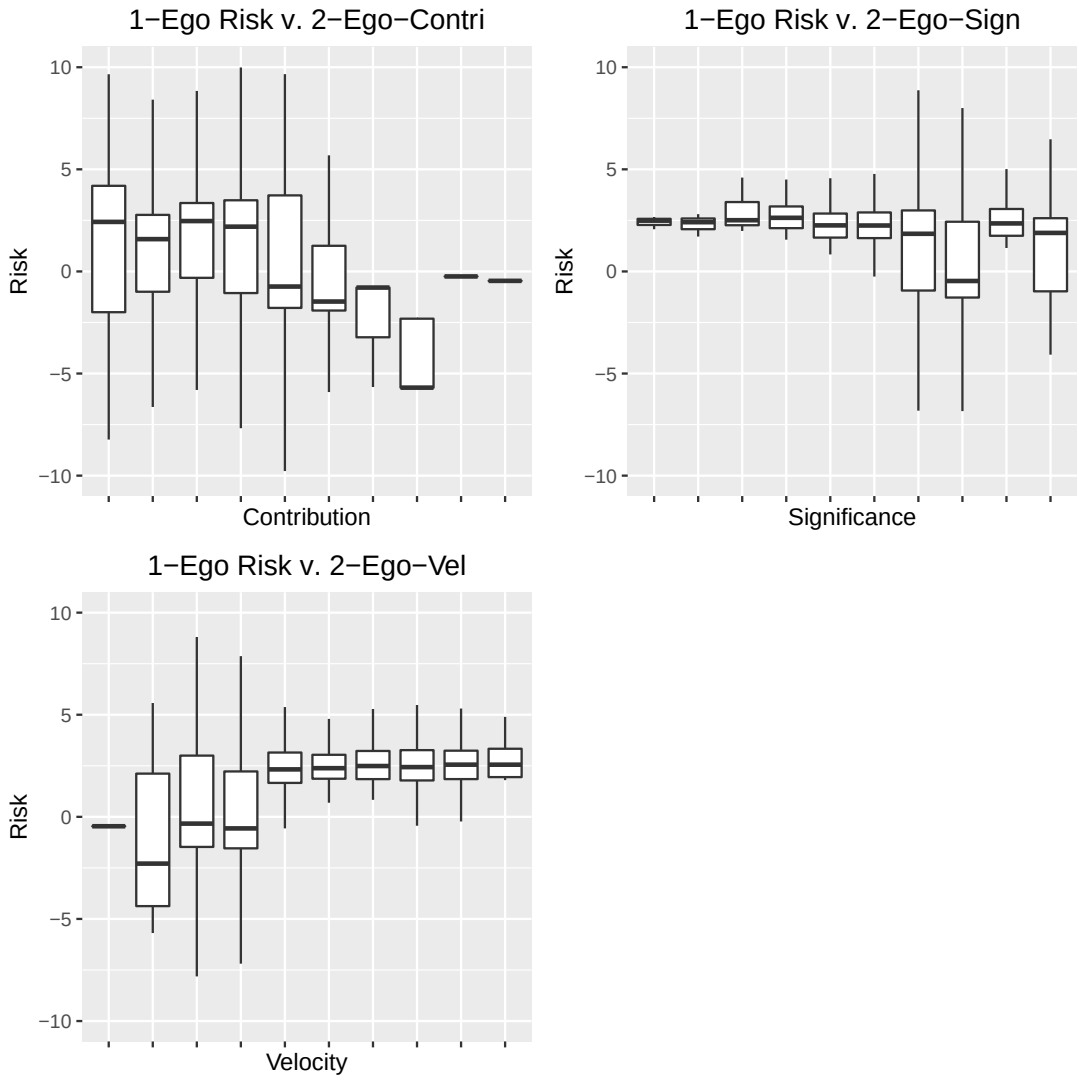


Figure 5.12: 1-Ego-Network Risk (CV of Velocity) plotted against 2-Ego-Network Risk Constructs

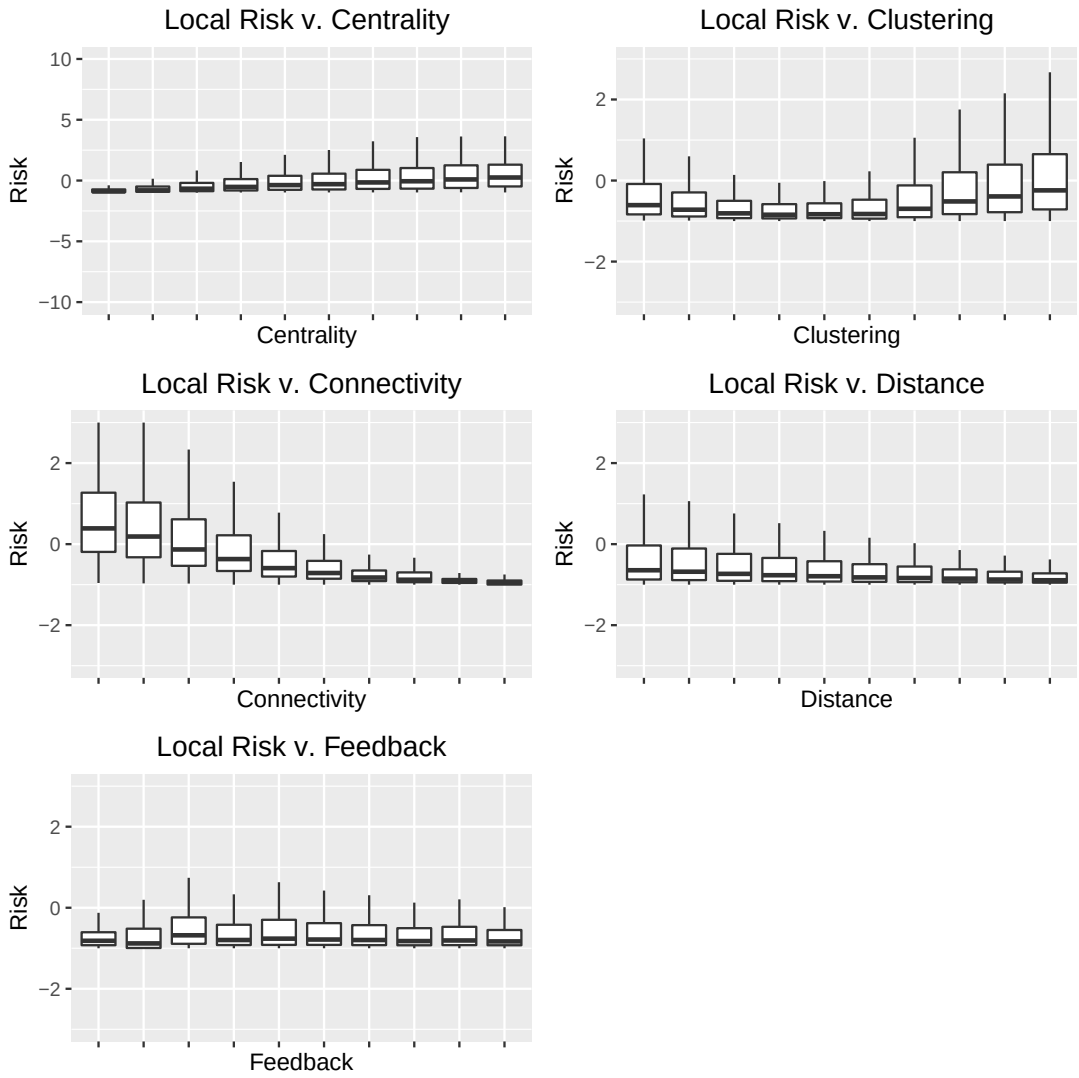


Figure 5.13: Local-Network Risk (CV of Velocity) plotted against Local-Network Structure Constructs

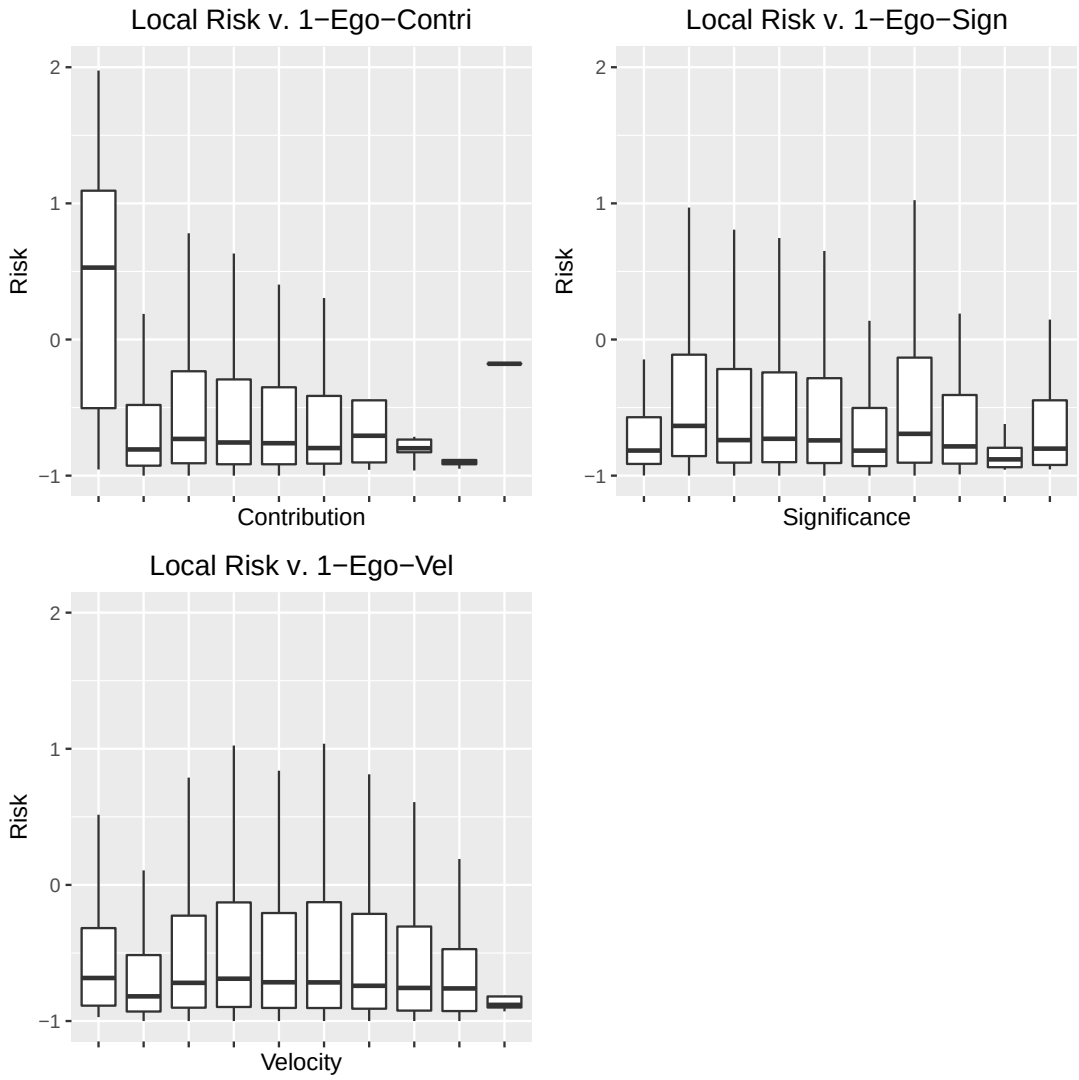


Figure 5.14: Local-Network Risk (CV of Velocity) plotted against 1-Ego-Network Risk Constructs

	Variables	VIF
1	alpha	1.67
2	beta	2.81
3	Centrality	6.17
4	Distance	5.08
5	Clustering	2.11
6	Feedback	2.66
7	Connectivity	2.39
8	avg_elrcf	45.32
9	avg_elrsf	7.70
10	avg_elrv	56.86
11	sd_elrcf	68.64
12	sd_elrsf	10.72
13	sd_elrv	61.40
14	cv_elrcf	4.69
15	cv_elrsf	11.57
16	cv_elrv	1.18
17	Centrality2	3.88
18	Distance2	4.56
19	Clustering2	1.52
20	Feedback2	1.49
21	Connectivity2	1.69

Table 5.21: The VIFs for the variables in the 2-Ego-Network Model.

.

concern. Further, all variables had a VIF of less than 10, once again indicating that fitting the linear models to these regressors would not pose multicollinearity concerns in the final model fits.

5.5 Results

5.5.1 2-Ego-Network Risk

The overall model fit of 8 out of the 9 equations in the 2-Ego-Network system had an R^2 of over 0.22, with the standard deviation of the risk significance being the best fit, as judged by R^2, at 0.66. A Wald χ^2-Test was conducted on the entire system with the restriction of $\beta_{i_1,j_1} = 0$. The result was significant ($p < 0.05$), indicating an overall good fit of the system. In addition, a second restriction χ^2-Test was conducted with the restriction of the parameters on the squared terms individually set to 0. Again, this test had indicated that the curvi-linear model is indeed a proper fit to the data

	Variables	VIF
1	alpha	1.67
2	beta	2.74
3	contribution	1.85
4	significance	1.90
5	velocity	1.89
6	Centrality	9.29
7	Distance	3.39
8	Clustering	3.98
9	Feedback	2.22
10	Connectivity	2.21
11	avg_elrcf	33.82
12	avg_elrsf	7.46
13	avg_elrv	44.12
14	sd_elrcf	45.24
15	sd_elrsf	10.76
16	sd_elrv	45.52
17	cv_elrcf	3.35
18	cv_elrsf	12.74
19	cv_elrv	1.16
20	Centrality2	8.37
21	Distance2	1.72
22	Clustering2	2.98
23	Feedback2	1.79
24	Connectivity2	1.88

Table 5.22: The VIFs for the variables in the 1-Ego-Network Model.

($p < 0.05$). In addition, most of the coefficients were found to be statistically significant ($p < 0.05$).

The effect that alpha and beta had on the respective dependent variables seemed to be fairly consistent in regards to their direction. Higher levels of the alpha parameter for the Beta distribution seemed to indicate higher levels of overall risk across most of the models. Higher levels of the beta parameter for the Beta distribution seems to portend lower levels of risk. In regards to the primary variable of interest in this dissertation, the Coefficient of Variation for the Risk Velocity, nearly all coefficients were found significant, although the R^2 was very low at a value of 0.04.

The resulting model indicates that there is a positive curvi-linear association between Distance and the Coefficient of Variation of Velocity (heretofore known as "Risk") and both coefficients were found significant ($p < 0.05$). This confirms Hypothesis 6. In

	Variables	VIF
1	alpha	1.33
2	beta	1.74
3	isNode	2.76
4	contribution	2.03
5	significance	1.20
6	velocity	1.70
7	Centrality	7.40
8	Distance	1.76
9	Clustering	2.52
10	Feedback	3.56
11	Connectivity	3.03
12	elrsf	2.29
13	elrcf	3.91
14	elrv	3.88
15	Centrality2	4.45
16	Distance2	1.23
17	Clustering2	1.45
18	Feedback2	1.44
19	Connectivity2	1.35

Table 5.23: The VIFs for the variables in the Local-Network Model.

addition, connectedness was found to be in a positive curvi-linear association($p < 0.05$), thus supporting Hypothesis 7. Feedback was also found to be significant ($p < 0.05$) and in a negative curvi-linear association, thus confirming Hypothesis 8. Centrality was found significant ($p < 0.05$) and in a negative curvi-linear relationship, thus confirming Hypothesis 9. Last, the model reveals that there is a statistically significant association between Clustering and Risk($p < 0.05$), and it was found to be negative curvi-linear, thus confirming Hypothesis 10.

5.5.2 1-Ego-Network Risk

The 1-Ego-Network results also had similar R^2 values to those of the 2-Ego-Network. Most values were above 0.22, however, the variables of interest only had an R^2 of 0.06. Overall, the fit of the model seems to be justified. I also ran a restricted Wald χ^2-Test with all of the parameters set equal to 0. The test confirmed that the model is indeed a good fit with the data ($p < 0.05$). A second restricted Wald χ^2-Test was conducted on only the parameters of the squared terms. Again the test indicated that the system

	μ_c	μ_s	μ_v	σ_c	σ_s	σ_v	c_c	c_s	c_v
(Intercept)	8.21(0.07)*	13.28(0.14)*	1.49(0.01)*	6.17(0.03)*	1.57(0.01)*	0.28(0)*	2.25(0.05)*	7.16(0.13)*	2.01(0.06)*
alpha	1.26(0.02)*	2.11(0.03)*	0(0)	0.94(0.01)*	0.24(0)*	0(0)*	0.09(0.01)*	0.51(0.03)*	0.02(0.01)
beta	-0.77(0.01)*	-1.24(0.01)*	0(0)*	-0.63(0)*	-0.17(0)*	0(0)*	0.03(0)*	-0.14(0.01)*	-0.04(0.01)*
Centrality	3.68(0.03)*	8.53(0.07)*	0.29(0)*	-0.35(0.01)*	0.36(0)*	0.09(0)*	2.7(0.03)*	7.52(0.06)*	1.09(0.03)*
Distance	2(0.05)*	4(0.1)*	0.12(0)*	0.34(0.02)*	0.32(0)*	0.05(0)*	1.07(0.04)*	2.96(0.09)*	0.71(0.04)*
Clustering	1.62(0.03)*	2.82(0.06)*	0.04(0)*	0.68(0.01)*	0.46(0)*	0.05(0)*	0.48(0.02)*	1.29(0.05)*	0.35(0.03)*
Feedback	1.23(0.03)*	2.09(0.06)*	0.01(0)*	1.23(0.01)*	0.48(0)*	0.03(0)*	0.04(0.02)	0.36(0.06)*	-0.07(0.03)*
Connectivity	0.57(0.06)*	0.79(0.11)*	-0.01(0)*	0.36(0.02)*	0.18(0)*	0.02(0)*	0.14(0.04)*	0.33(0.1)*	-0.01(0.05)
Centrality2	-0.45(0.01)*	-1(0.02)*	-0.04(0)*	0.08(0)*	-0.02(0)*	-0.01(0)*	-0.3(0.01)*	-0.72(0.02)*	-0.24(0.01)*
Distance2	-0.39(0.03)*	-0.56(0.06)*	0.01(0)*	0(0.01)	-0.02(0)*	-0.01(0)*	-0.27(0.02)*	-0.54(0.06)*	0.08(0.03)*
Clustering2	-0.02(0.02)	0.04(0.05)	0(0)*	0.01(0.01)	0.02(0)*	0(0)*	0.04(0.02)*	0.12(0.04)*	-0.08(0.02)*
Feedback2	0.06(0.03)*	0.36(0.05)*	0.02(0)*	0.18(0.01)*	0.06(0)*	0(0)	-0.03(0.02)	0.11(0.05)*	-0.1(0.02)*
Connectivity2	0.67(0.07)*	0.92(0.13)*	0.02(0.01)*	-0.49(0.03)*	-0.12(0.01)*	0(0)*	0.6(0.05)*	1.13(0.12)*	0.39(0.06)*
σ	5.13	9.94	0.4	1.96	0.45	0.1	3.87	9.28	4.64
R^2	0.34	0.38	0.26	0.51	0.67	0.48	0.28	0.35	0.04
Adjusted R^2	0.34	0.38	0.26	0.51	0.67	0.48	0.28	0.35	0.04

Table 5.24: Results from SUR Estimation of 2-Ego-Network System

	μ_c	μ_s	μ_v	σ_c	σ_s	σ_v	c_c	c_s	c_v
(Intercept)	9.45(0.08)*	13.76(0.12)*	1.38(0.01)*	6.11(0.03)*	1.43(0.01)*	0.27(0)*	3.04(0.04)*	7.79(0.08)*	1.68(0.07)*
alpha	0.81(0.01)*	0.95(0.02)*	-0.01(0)*	0.93(0.01)*	0.19(0)*	0(0)*	0(0.01)	0.07(0.02)*	-0.06(0.01)*
beta	-0.5(0.01)*	-0.51(0.01)*	0.01(0)*	-0.64(0)*	-0.13(0)*	0(0)*	0.05(0)*	0.06(0.01)*	0.04(0.01)*
Centrality	6.88(0.24)*	14.08(0.35)*	0.5(0.02)*	-1.67(0.09)*	0.21(0.02)*	0.12(0)*	5.26(0.13)*	11.78(0.24)*	1.13(0.2)*
Distance	2.53(0.05)*	4.5(0.07)*	0.14(0)*	0.24(0.02)*	0.27(0)*	0.04(0)*	1.27(0.03)*	2.98(0.05)*	0.63(0.04)*
Clustering	2.43(0.04)*	3.73(0.06)*	0.12(0)*	0.56(0.02)*	0.49(0)*	0.06(0)*	0.75(0.02)*	1.61(0.04)*	0.63(0.03)*
Feedback	0.92(0.03)*	1.21(0.04)*	0(0)*	0.95(0.01)*	0.33(0)*	0.02(0)*	0.05(0.01)*	0.21(0.03)*	-0.05(0.02)*
Connectivity	0.08(0.03)*	-0.02(0.04)	-0.01(0)*	0.4(0.01)*	0.14(0)*	0.01(0)*	-0.24(0.02)*	-0.47(0.03)*	-0.04(0.03)
contribution	0.09(0)*	0.18(0)*	0.01(0)*	0.03(0)*	0.01(0)*	0(0)*	-0.01(0)*	-0.01(0)*	0.01(0)*
significance	-0.25(0.01)*	-0.52(0.01)*	-0.03(0)*	0.13(0)*	-0.01(0)*	-0.02(0)*	-0.39(0)*	-0.75(0.01)*	-0.15(0.01)*
velocity	0.23(0)*	0.4(0)*	0.01(0)*	-0.17(0)*	0(0)	0.01(0)*	0.23(0)*	0.45(0)*	0.1(0)*
Centrality2	5.31(0.33)*	9.29(0.49)*	0.32(0.02)*	-1.94(0.13)*	0.09(0.03)*	0.1(0)*	3.78(0.18)*	7.34(0.34)*	0.89(0.28)*
Distance2	-0.1(0.06)	-0.05(0.08)	0.01(0)*	-0.02(0.02)	0.05(0)*	0.01(0)*	-0.06(0.03)	-0.09(0.06)	0.11(0.05)*
Clustering2	-0.32(0.02)*	-0.51(0.03)*	-0.03(0)*	0.06(0.01)*	0(0)	0(0)*	-0.15(0.01)*	-0.28(0.02)*	-0.18(0.02)*
Feedback2	-0.15(0.02)*	-0.07(0.04)	-0.01(0)*	0.34(0.01)*	0.06(0)*	0(0)*	-0.13(0.01)*	-0.12(0.02)*	-0.2(0.02)*
Connectivity2	-0.01(0.02)	0.04(0.03)	0(0)*	-0.09(0.01)*	-0.05(0)*	0(0)*	0.06(0.01)*	0.13(0.02)*	0(0.02)
σ	5.12	7.54	0.35	2.05	0.39	0.07	2.82	5.29	4.32
R^2	0.36	0.4	0.22	0.57	0.66	0.58	0.37	0.4	0.06
Adjusted R^2	0.36	0.4	0.22	0.57	0.66	0.58	0.37	0.4	0.06

Table 5.25: Results from SUR Estimation of 1-Ego-Network System

model is indeed a proper fit ($p < 0.05$).

The effect that alpha and beta had on the risk were analogous to their effect on the 2-Ego-Network risk. Alpha was found to be positively associated with risk while Beta was the opposite. Connectivity was found to be neither significant in the linear term nor significant in the quadratic term ($p > 0.05$). This unfortunately will not support Hypothesis 16. Interestingly, Centrality was found to be in a positive curvi-linear relationship with risk (ego1-risk) ($p < 0.05$), which is contrary to the results from the 2-Ego-Network model. Unfortunately, however, Hypothesis 12 would be invalidated by this interesting observation. The Clustering construct was found to be in a negative association with risk, thus confirming Hypothesis 13 ($p < 0.05$). Feedback was found to be statistically significant($p < 0.05$), and it did exhibit a negative curvi-linear pattern, thus confirming Hypothesis 14. Furthermore, distance was found to be statistically

	Contribution	Significance	Velocity
(Intercept)	2.2(0.03)*	5.18(0.03)*	-0.43(0.01)*
alpha	0.38(0)*	0.95(0)*	-0.02(0)*
beta	-0.28(0)*	-0.63(0)*	0.01(0)*
isNode	1.02(0.02)*	2.25(0.02)*	-0.04(0)*
Centrality	0.39(0.02)*	-1.91(0.03)*	0.29(0.01)*
Distance	-0.28(0.01)*	1.04(0.01)*	-0.18(0)*
Clustering	-0.04(0.01)*	-1.46(0.01)*	0.14(0)*
Feedback	0.27(0.01)*	1.55(0.01)*	-0.09(0)*
Connectivity	-1.87(0.01)*	-0.24(0.01)*	-0.42(0)*
contribution	0.02(0)*	0.03(0)*	0(0)*
significance	0.01(0)*	0.04(0)*	-0.02(0)*
velocity	0(0)*	-0.04(0)*	0.01(0)*
Centrality2	0.47(0.02)*	0.92(0.02)*	0.02(0)*
Distance2	-0.02(0.01)*	0.12(0.01)*	0.01(0)*
Clustering2	0.18(0.01)*	0.32(0.01)*	0(0)
Feedback2	-0.27(0.01)*	-0.3(0.01)*	-0.03(0)*
Connectivity2	0.68(0.01)*	0.04(0.01)*	0.16(0)*
σ	2.37	2.57	0.61
R^2	0.41	0.46	0.36
Adjusted R^2	0.41	0.46	0.36

Table 5.26: Results from SUR Estimation of Node/Edge System

significant ($p < 0.05$) and had a positive curvi-linear association. This would indeed validate Hypothesis 15. Last, connectedness was found to be statistically insignificant, hence, Hypothesis 16 could not be validated.

5.5.3 Local Risk

The last data set comprised of the node and edge level data. The model was ran, but different from the prior two systems. This system did indeed still have the three measures of risk for 1-Ego-Network as the independent variables, respectively, but it also only had three dependent variables. The overall fit of the model seemed to be better than the prior two systems. Indeed, the smallest R^2 value was only 0.36. Overall most statistics were shown to be significant. Again, a Wald χ^2-Test was conducted on the parameters with the restriction of them set all to 0 individually. The results showed that the model was an overall good fit ($p < 0.05$). A second restriction test was conducted on the parameters but only those of the squared terms. Again, the Wald Test revealed that the model was an overall good fit to the data ($p < 0.05$).

The three control variables of alpha, beta, and isNode, were all found to be statistically significant ($p < 0.05$). Alpha and beta once again was shown to have a pattern consistent across the all the models. The results indicate that there is a statistically significant relationship between Centrality and Risk. This was found to be significant at ($p < 0.05$) and indicates that the association is positive curvi-linear, hence invalidating Hypothesis 17. Clustering was not found to be statistically significant at this level of analysis ($p > 0.05$), and hence, Hypothesis 18 cannot be properly justified from this research. Feedback was found to be in a negative association with risk ($p < 0.05$), and transcends a negative curvi-linear relationship, which is in line to my Hypothesis 19. Distance was found to be statistically significant, and in a positive curvilinear trend ($p < 0.05$), consistent with Hypothesis 20. Last, Connectedness was found to be a positive curvilinear pattern ($p < 0.05$), hence, Hypothesis 21 is supported.

5.5.4 Cross-Risk and Structural Differences

The last part of the analysis was to conduct a series of structural tests to determine if the various structural properties change from the 2-Ego-Network to the lower levels, and so on. A simple comparison of means between two populations was conducted. The results are summarized in Table 5.3. As we can see, the 2-Ego-Network Distance tends to be greater than that of the 1-Ego-Network, which would confirm Hypothesis 1($p < 0.05$). Next, we observe that the Centrality in the 2-Ego-Network is larger than the Centrality in the 1-Ego-Network, which would indeed confirm Hypothesis 2($p < 0.05$). For the last three structural characteristics, namely Clustering, Feedback, and Connectivity, we notice that the Ego-Network actually held larger values in these variables than it's 2-Ego-Network. Hypotheses 3 and 5 were hence confirmed($p < 0.05$), but unfortunately Hypothesis 4 was not not. Last, we notice that the velocity (risk of the network) is larger for the 2-ego-network than it is for the ego-network ($p < 0.05$). Hence, Hypothesis 11 is supported as well.

While I have not argued or designed any hypotheses regarding the structural differences between the 1-Ego-Network and the Local Network, I would like to illustrate some of the comparisons drawn from these two data sets. First, we notice that the

Centrality of the Ego-Network is actually less than that of the Local element. This makes sense, as I had explained earlier, the local elements are actually based on global network properties. Next, we notice that the distance, clustering and feedback are all less in the Local element than within the Ego-Network. Finally we notice that the risk in the ego network tends to be higher than the risk at the node level. If we turn to Table 5.3, we will see a comparison drawn between the nodes and edges. We notice that the centrality for the edges is higher than that of the nodes. Again, this is in line with our earlier reasoning. Next, the Distance and Clustering are actually greater at the node level rather than at the edge level. Feedback, however, is less in the nodes than in the edges. Last, the connectivity tends to best higher in the edges than in the nodes. The risks of the edges, however, are generally higher than the risk of the nodes.

Looking back at the Local SUR Model, we can see that the coefficient on the variable isNode is negative and significant ($p < 0.05$). Therefore, Hypothesis 22 is supported, and we would conclude that Edge structure has a larger effect on edge risk than node structure has on node risk. In addition, we do see, however, that the 2-Ego-Network risk has a positive linear effect on the 1-Ego-Network risk ($p < 0.05$), and hence, Hypothesis 24 is supported. The same observation is made regarding the effect that the 1-Ego-Network risk has on the Local Node/Edge Risk ($p < 0.05$). Therefore, Hypothesis 25 is also supported.

5.6 Summary of Hypothesis Test Results

Hypothesis Number	Description of Hypothesis	Result
1	2-Ego-Network Distance > 1-Ego-Network Distance	Supported
2	2-Ego-Network Centrality > 1-Ego-Network Centrality	Supported
3	2-Ego-Network Clustering < 1-Ego-Network Clustering	Supported
4	2-Ego-Network Feedback > 1-Ego-Network Feedback	Not Supported
5	2-Ego-Network Connectivity < 1-Ego-Network Connectivity	Supported
6	2EN: Distance +CL Association with 2EN Risk	Supported
7	2EN: Connectedness +CL Association with 2EN Risk	Supported
8	2EN: Feedback -CL Association with 2EN Risk	Supported
9	2EN: Centrality -CL Association with 2EN Risk	Supported
10	2EN: Clustering -CL Association with 2EN Risk	Supported
11	2-Ego-Network Risk > 1-Ego-Network Risk	Supported
12	1EN: Centrality -CL Association with 1EN Risk	Not Supported
13	1EN: Clustering -CL Association with 1EN Risk	Supported
14	1EN: Feedback -CL Association with 1EN Risk	Supported
15	1EN: Distance +CL Association with 1 EN Risk	Supported
16	1EN: Connectedness +CL Association with 1EN Risk	Not Supported
17	L: Centrality -CL Association with L Risk	Not Supported
18	L: Clustering -CL Association with L Risk	Not Supported
19	L: Feedback -CL Association with L Risk	Supported
20	L: Distance +CL Association with L Risk	Supported
21	L: Connectedness +CL Association with L Risk	Supported
22	Edge Structure on Edge Risk > Node Structure on Node Risk	Supported
23	Risk Trickles Down k-Ego-Networks Sequentially	Future Research Topic
24	2 Ego Network Risk +L Association with 1 Ego Network Risk	Supported
25	1 Ego-Network Risk +L Association with Local Network Risk	Supported

Table 5.27: The set of hypotheses and their results.

Chapter 6

Discussion, Limitations and Managerial Implications

6.1 Discussion of Results

The results of the data analysis offer some interesting insights into how the structure of a supply network may affect the risk and the risk propagation of the network. Such results can be used to isolate a hidden conceptual framework of risk propagation. From the 2-Ego-Network level, I have demonstrated that it's structure is inherently different than than the 1-Ego-Network level. Further, I argued theoretically that as firms change and manipulate their 1-Ego-Networks, a chain reaction occurs, which may result in a very drastic change of the firm's 2-Ego-Network. Another aspect of the results is that the structure of the various ego networks and local node/edges impact risk based on how they are situated with respect to the 5 Fundamental Structural dimensions. I also have demonstrated that risk of which is inherent within the 2-Ego-Network is different than the risk that is inherent within the 1-Ego-Network (namely, its larger). In addition, I have demonstrated that the risk within the 1-Ego-Network is larger than the risk within the local node/edge. And subsequently, I have demonstrated that the risk of edges is often higher than the risk of nodes.

That is, I have empirical evidence to suggest some initial observations that can be used to construct a generic conceptual model of risk propagation. From what I have verified in this dissertation, I have shown that 2-Ego-Network Risk > 1-Ego-Network Risk > Edge-Risk > Node Risk. This ordering of structures in the network allows the theorist to better understand how risk "propagates" to a focal firm, outward-to-inwards. I would imagine that it is not too difficult to infer that given a node n in the network, such that it's highest k-ego-network level is m (that is, n's m-ego-network = the entire supply network), that risk has the tendency to emanate from the outside and

downwards through the various k-ego networks. That is, risk begins at m, trickles down to $(m-1)$, etc. As I have demonstrated, we have an ordering in place, that essentially states that risk will decrease or reduce as it moves down the various k-ego-networks. I have termed this theorem the "The Theorem of Relativity for Supply Chain Risk Propagation", since this is paradoxically a network phenomena, but the flow is to be understood from the perspective of multi-level embedded structures, of which are all *relative* to a specific focal firm. This sequence of structures is specific to the firm, not the network, despite it characterizing movement within the network itself. Of course this study is not sufficient proof for this phenomena, but it is an interesting observation that the results seem to imply towards a transcendence of a more conceptual framework of understanding and managing supply chain risk propagation.

Other than the potential conceptual motivations that the results of the analysis seem to infer, there are strategic and practical understandings that can be extracted from the results. I have demonstrated that structural characteristics of the embedded structures affect risk differently. For example, the results indicate that the Centrality construct, at the 2-Ego Network Level, will be related to risk in a negative manner. Therefore, firms within a firm's 2-Ego Network may strive to increase the entire group's centrality as a means to reduce overall 2-Ego-Network risk. What is interesting is that there is a paradoxical effect at the Ego-Network level. Namely, when firms within the ego-network increase their centrality, they increase the ego-network's risk. The same holds true at the node and arc level, according to the results. This implies that a struggle is at play. Which strategy should a firm pursue? Encouraging their neighbors and their neighbor's neighbors to increase centrality levels for the "greater good", or, should the firms and ego-networks be more self-fulfilling, and decrease their centrality levels for their benefit? This remains an observation that will be left to answer for future research.

The results indicate that centrality is really the only metric that exhibits this characteristic of opposing effects on risk. All the other dimensions seem to be fairly aligned in how they respectively affect the respective structure's risk. However, while structurally they affect risk in the same manner (either negative curvilinear or positive curvilinear),

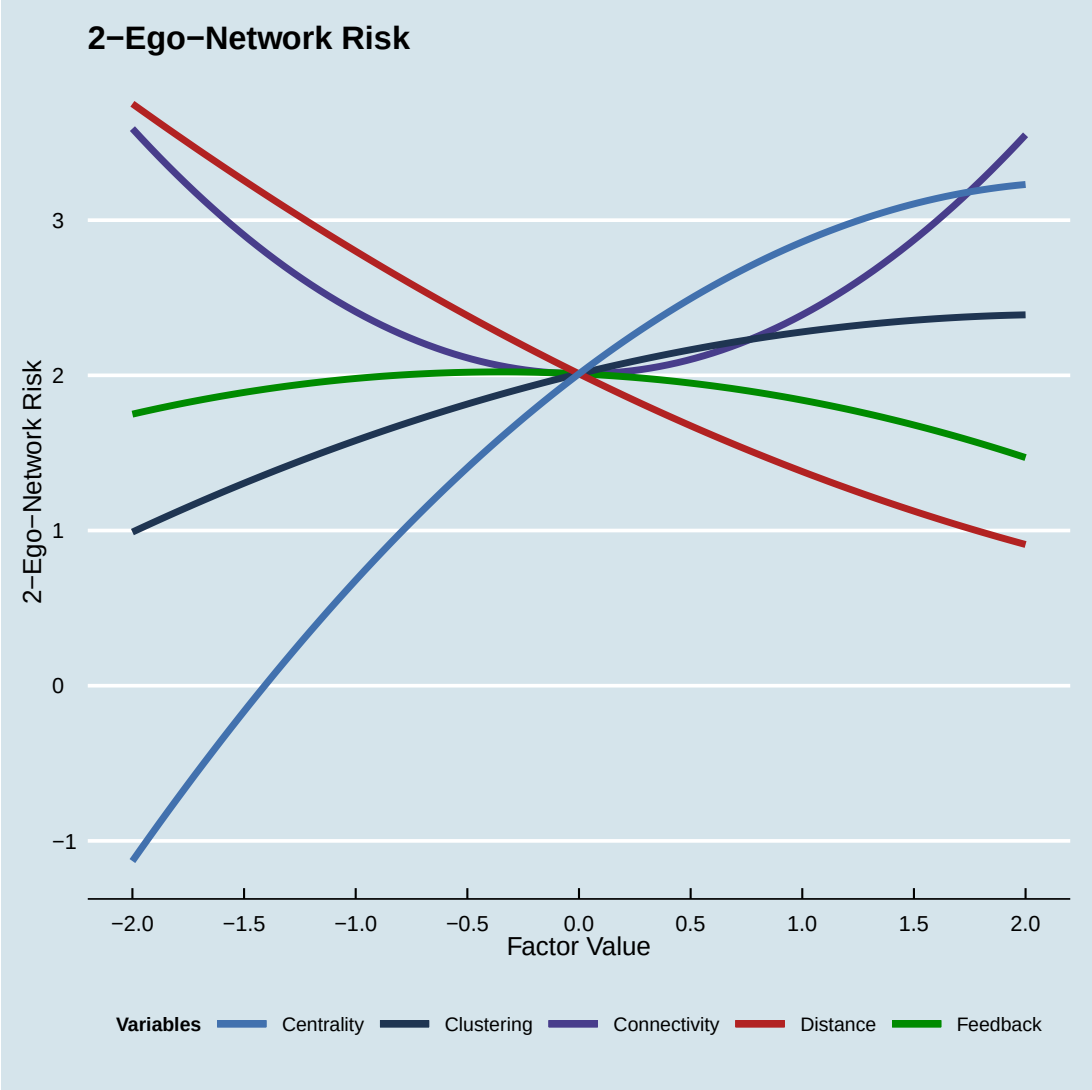


Figure 6.1: 2-Ego Network Risk Structural Model

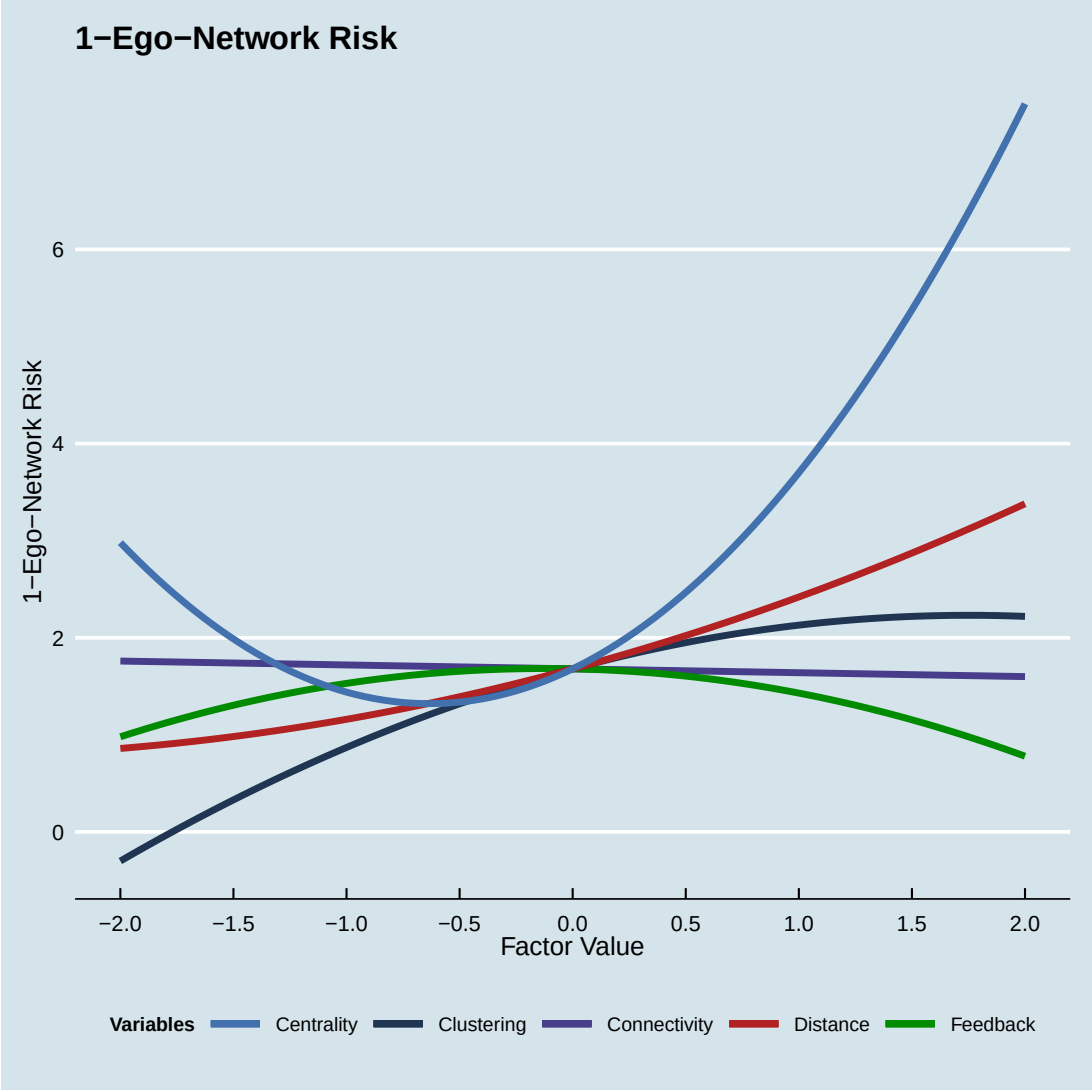


Figure 6.2: 1-Ego Network Structural Model

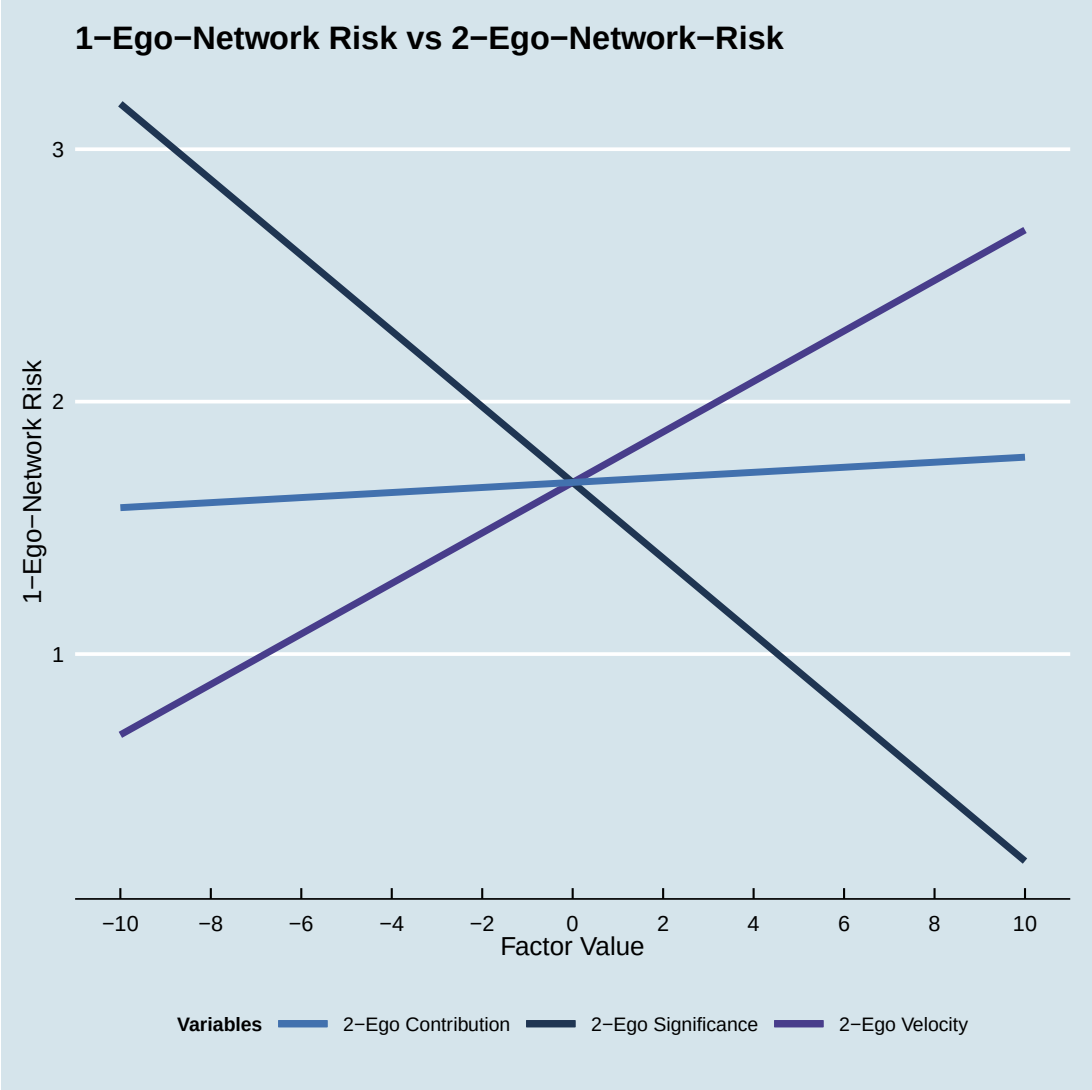


Figure 6.3: 1-Ego Network Risk Model

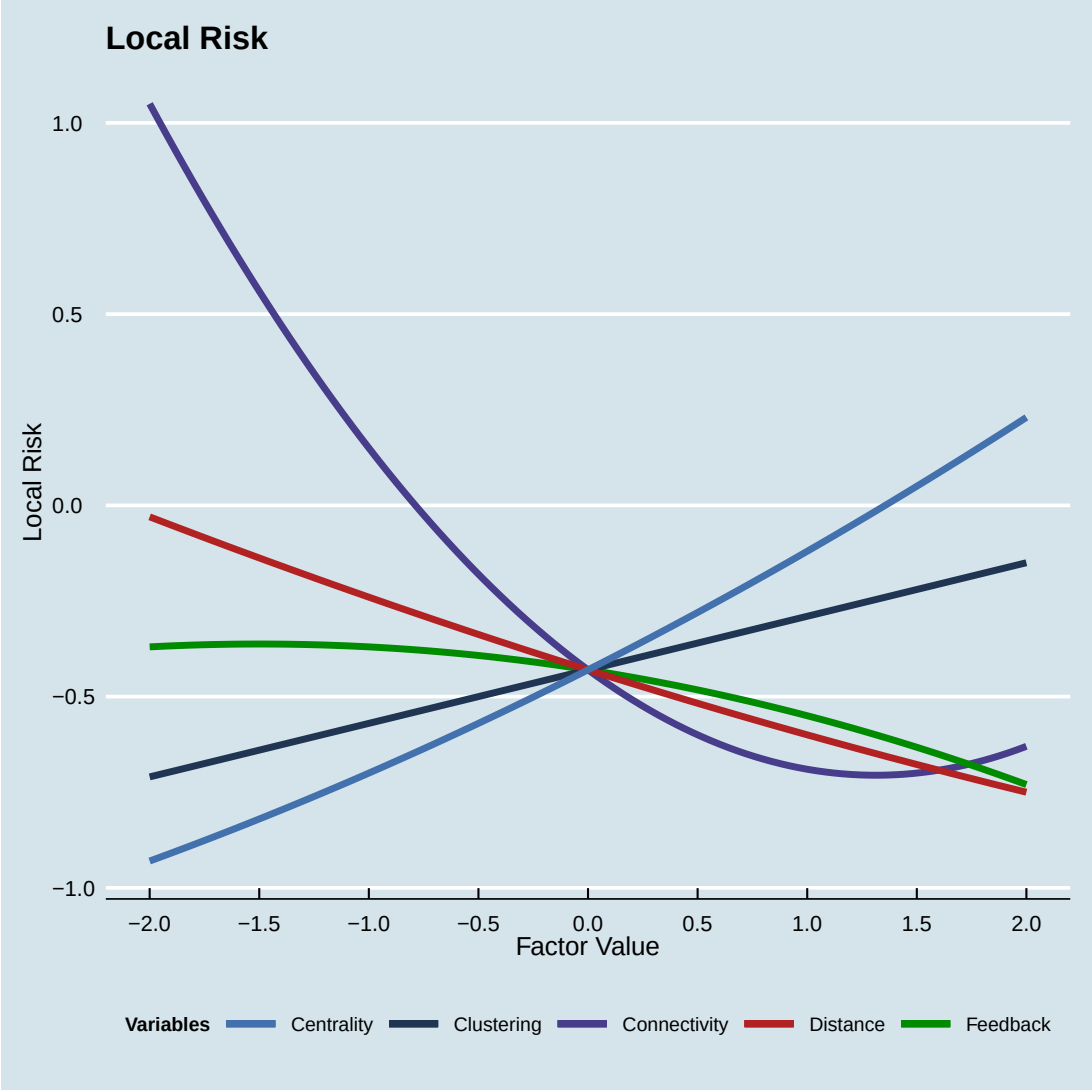


Figure 6.4: Local Network Structural Model

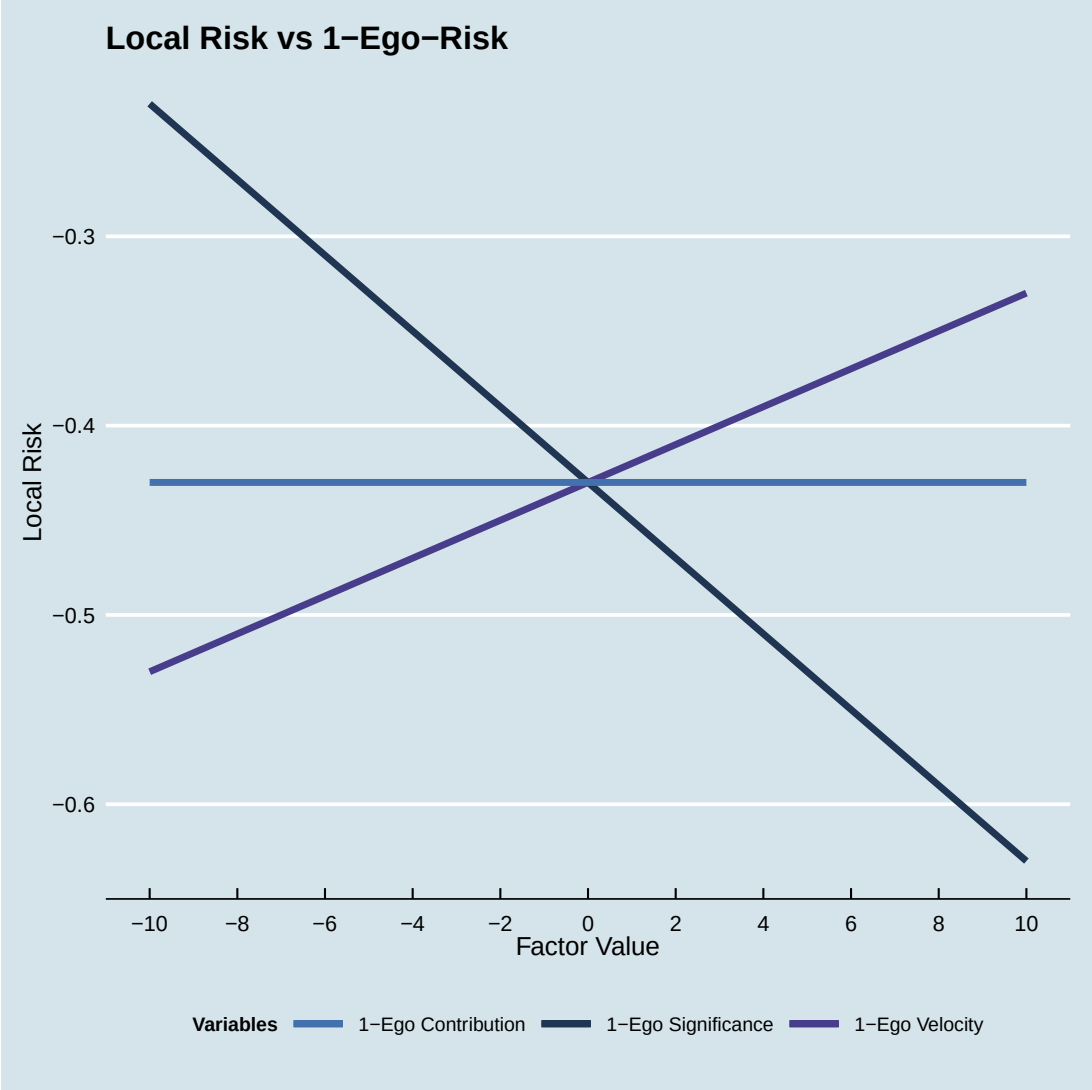


Figure 6.5: Local Network Risk Model

the shapes of these curves may very well differ from level to level. As these are curvilinear, there is always a trade-off to consider. For example, in the case of a dimension being negatively curvilinear, the question for the firm is, should it reduce it's dimension (either cluster or connectedness, for example), in order to achieve lower risk levels, or, should they increase it well beyond the point where the maximum occurs? There may be benefits in doing so for either case, and of course would depend on the specific context at hand.

Other than these observations of the dynamics between the different levels of the k-ego-networks and risk, we can visualize the curvilinear effects that the models happen to suggest. Figures 6.5,6.4,6.3,6.2,6.1 illustrate the estimated curvilinear models on the various levels of risk. We can make a few observations, and identify potential trade-offs:

1. Increasing distance in the 2-Ego-Network will result in a decrease of risk. However, the same action in the 1-Ego-Network will actually increase risk. Increasing distance at the node level will decrease risk.

2. Increasing centrality at the 2-Ego-Network level will increase risk. However, this risk appears to either reach a carrying capacity or reverse direction entirely. The carrying capacity is the more plausible speculation. This essentially says that the 2-Ego-Network's risk is capped with respect to centrality. Once the 2-Ego-Network reaches a certain centrality level, then the marginal effects on risk for marginal additions to centrality is practically nothing.

3. At the 1-Ego-Network level, however, we notice that increased centrality, initially, will result in a reduction of risk. However, once it reaches past a certain point, it will accelerate. That is, the relationship between centrality and risk differs between the 2-Ego-Network and the 1-Ego-Network.

4. We notice that local centrality is the worst offender. There is no gain in regards to risk by increasing centrality locally, since doing so will always result in an increase in risk Amazingly, this would partially contradict Choi's speculations on the relationship between complexity and risk. Recall that he argues firms that increase their inter-relationships should experience lower levels of risk. This may

be true for inner-firm, or single location risk, but as these results illustrate, risk propagation itself would be accelerated with increased levels of inter-relationships. This indicates that firms need to be careful in weighing the benefits of reducing within-firm risk while increasing risk propagation as a result.

5. Clustering is the same across all dimensional levels. It appears to impact risk in similar ways that other dimensions do. Therefore, we can speculate that this is the most "stable" dimension. Furthermore, in the 2-Ego-Network level, it reaches a carrying capacity much soon than centrality. This would imply that 2-Ego-Networks are much more robust to risk when changes to clustering have ensued.

6. Connectivity appears to be a "wolf in sheep's clothing". Across nearly all the dimensions, a familiar pattern emerges. Once the connectivity increases, the entity enjoys an initial reduction in risk. However, if the entity becomes too-connected, risk appears to return. That is, it is a "consume in moderation" dimension for those who seek to minimize risk via connectivity.

7. Of all the dimensions, feedback appears to the dimension that offers the most amount of benefits. Increasing the feedback at any level of the network appears to always result in reduction of risk.

8. A pattern across the levels of analysis is how the risk "flows" to the lower levels of analysis. We can see from Figure 6.3 that as the risk significance increases in the higher dimension, the lower dimension will experience a drop in their own risk. We also notice that as risk velocity increases in the higher dimension, risk in the lower dimension also increases. Contribution seems to have little impact on lower dimensional risks.

6.2 Contributions of this Research

As far the author of this study is aware, this is the first major study to properly distinguish between supply chain disruption propagation and supply chain risk propagation. Many in the current literature have confounded these two constructs (Fan & Stevenson,

2018), however, while some have taken the more careful approach of distinguishing between supply chain risk and supply chain disruption (DuHadway et al., 2017), no one, to the best of this author's knowledge, have attempted to distinguish (1) between supply chain risk and supply chain risk propagation and (2) distinguish between supply chain risk propagation and supply chain disruption propagation. Furthermore, there has been a gap in the extant literature in attempting to explain why it is that just so many risk measurement frameworks exist, as well as classification systems. This research is the first of it's kind to provide a philosophical foundation to better understanding why so much diversity and disagreement within the risk literature is still salient to this day.

Another issue that this author has identified is a lack of study between the link of supply chain structure and a variety of supply chain constructs. I have explained why this gap exists: there are too many characterization frameworks for explaining and describing supply network structure. This dissertation has taken the time to review through many of these and to synthesize the literature into three main and primary areas of this research stream: Context-Dependent, Context-Free and Context-Dependent Classification, and Graph-Theoretical. Despite the plethora of methodologies that are salient in the extant literature to accomplish a proper understanding of supply network structure characterization, few attempts have been made to generalize it to a perspective where all other viewpoints can easily be incorporated. The conceptual characterization of general networks is the first of it's kind, as far as this author is aware. Given that I have isolated the 5-fundamental conceptual dimensions of a general network, this conceptual framework can now be applied to a variety of areas, not just limited to supply networks, to gain a more thorough and practical understanding of networks, without getting caught up in the vast plethora of graph-theoretic and social-theoretic metrics.

In addition, another contribution that this dissertation has made is within the supply chain visibility stream. I have taken great care to synthesize the horizon-based literature and fuse it with the structure-based literature, to design a new structure that is (1) practical, (2) easy to understand, (3) easily implemented into other conceptual and mathematical frameworks, and (4) can act as a measure of *supply chain visibility*. As (Carter et al., 2015) has mentioned, practically no research exists that study the

dynamics of a supply network outside the attenuated horizon (Carter et al., 2015). The construct I have introduced, the *k-ego-network*, allows the supply chain researcher a measure and method for better understanding this attenuated horizon, and I foresee it's usage for better understanding these dynamics that occur outside the horizon, yet, "trickle down" to within the horizon. The *k-ego-network* will now allow researchers a tool to understand more complicated constructs in a simple and elegant manner.

This research has also contributed to field of supply chain risk propagation, as the primary topic itself of course. Risk propagation research is still in it's infancy (Garvey et al., 2015), and the necessary theoretical and analytical constructs and tools, respectively, have yet to be matured. I have taken great care to accomplish both in this dissertation. I have introduced the general constructs of *contribution*, *significance*, and *velocity* in regards to risk propagation. Admittedly, these constructs initially began from a quest to measure risk. Ironically, the measures themselves I have managed to generalize to more practical abstract notions of "diagnostic", "causal", and "change", from the perspective of risk propagation. In addition to the new constructs, I have also introduced measure-based operationalizations of the constructs (conveniently named the same, and dependent on the general framework of measuring and understanding risk propagation, namely (Garvey et al., 2015) framework).

Another contribution to the literature was the formalization of a framework for the supply network level of analysis. While this conceptualization is nothing new to the literature, it did have gaps. The original conceptualization lost clarity of the levels of analysis the higher up dimensionally they went (Peck, 2005). In addition, the original framework for understanding the different levels of analysis of a supply network was not considering one of the most important actors in the network: the consumer. I have extended this conceptual framework of supply network level dimensions to include two higher-order dimensions and one extreme low dimension (i.e. the people-supply-network). I subsequently have leveraged this framework to explain and justify why the current applications of reliability-based theory, disruption analysis, and fault-tree and Petri-Net models are not a proper application for studying supply chain risk and supply chain risk propagation.

As I argued earlier, many authors in the extant literature think in a singular and linear dimension when it comes to disruption and "risk" (many still conflate these two) analysis. Nearly all models have assumed that once a disruption "stops" at a location, it is absolutely impossible for it to continue to "trickle down". I have argued, leveraging the construct of the "people-supply-network", that this is an unrealistic and misguided assumption, something that the extant literature has been exercising for years. The many models of fault-analysis and more deterministic approaches have entirely ignored certain aspects of the contextual constructs that are necessary to understand in just so how a disruption might "continue", despite it paradoxically being "stopped". I have offered a very convincing argument above, and continue to argue that the only approach to achieve balance between modeling efficiencies versus realities is to precisely use a Bayesian Network approach. I argued in this dissertation that such an approach allows the researcher and practitioner to account for "unanticipated risk", which is something that nearly all the deterministic models miss.

Last, I have contributed to an area of research that is very scarce of publication, but has much opportunity to grow. The relationship between supply network structure and supply network risk propagation has been proposed and empirically validated by this research. The fundamental 5-dimensions of any network have been related to risk, and a greater understanding of how this interaction between structure and risk play out together will certainly go a long way in providing a foundation for future researchers to build upon, of which will better aid researchers to understand and craft proper mitigation strategies that not only consider contextual factors during the decision making process, but technical factors, but from an easy to understand and simplistic conceptual framework.

In addition, to the best of this author's knowledge, no attempt has been made to understand the dynamics between supply chain ego-network structure and risk. This is actually quite surprising, since the incorporation of ego-networks into the supply chain literature has now spanned well over two decades. Yet another interesting gap in the literature was to understand how risk affects nodes and arcs, and if they do so differently. In this research I have suggested and verified that arc-based risk and

node-based risk is indeed inherently different. If not due to contextual reasons, then certainly, as I have demonstrated in this dissertation, due to pure structural reasons. I also have demonstrated that ego-network structure differs at different level of k-ego-networks, and have argued that the inherent structural characteristics of ego-networks are different than those of individual nodes or arcs,despite them together comprising of an ego network or a higher-order ego-network.

6.3 Limitations of Research

While I remain confident in the results of this study, no research is without it's potential sources of problems to creep in and disrupt the ability to generalize any results. My study is of course no different. While I have taken the utmost careful approach to ensure that every step in this research was conducted thoroughly and without bias, there are limitations in the ability to generalize these results.

First, the networks I had generated were not based on an evolutionary algorithm that would result to "real world networks". In addition, they certainly were not based on real-world supply network data. I made careful assumptions regarding how supply networks form, but not nearly as restrictive as others in the literature have, especially within the complex systems and evolutionary network streams of literature. Hence, the external validity of the results of this study may indeed be at risk. With that said, I had attempted to achieve as much heterogeneity within the dataset so as to properly isolate the effects of structure on risk.

Next, the framework for risk measurement had assumed that a "consequence" was defined as the number of nodes that either were affected or could be affected due to a disruption in the network. The results of this study may very well change under the assumption that "consequence" is not measured as a form of count, but rather of monetary value, or some other form of utility. Therefore, the results of this study can only be argued to hold within environments where actors consider "consequence" from the perspective of "number of areas affected".

The simulation of this study was designed to randomly sample conditional probabilities from a Beta Distribution of a given α and β parameter. Hence, the generalizability of these results would only logically hold within environments where all the conditional probabilities within a network themselves follow a Beta Distribution. This unfortunately is a restriction to the current study at hand, as it may well be the case that probability and structure are interlinked, or, that probabilities of disruptions may follow certain patterns in specific areas of a supply network, which is a characteristic that the Beta Distribution would most likely not be able to reflect.

Yet another limitation of this study is the lack of consideration of time. Disruptions and risk always have delays and the dimension of time involved. This study had sought to understand the connection between structure and risk in a static sense. However, natural supply networks do not operate in a static world, and, as such, the networks are ever changing. Therefore, consequentially, the results of this study can only be spoken from a single moment in time. Further study would need to be conducted to understand the evolutionary dimensions and attributes of the relationships between structure and risk.

I also argue that another limitation of this study is within some of the measures I had employed for the construct of "Feedback". While I firmly trust the results of this study, and proper variables appeared to have loaded on the construct properly, there are other measurements I could have employed that would, I argue, have better reflected this construct. Measurements such as graph circumference, average cycle size, and total number of cycles as well as total number of largest/smallest cycles would have offered additional clarity to this construct. Unfortunately, due to time constraints and a lack of proper algorithms to identify these fully, I made the difficult decision to ignore these observables. While I feel this is a limitation, I am confident that a follow up study could employ these measures and should be able to draw similar conclusions as I have here.

Last, a limitation of this study is due to the calculation of the risk measures themselves. I have demonstrated, despite past criticism from some in the extant literature, that it is possible to conduct inference within large Bayesian Networks in an efficient

manner, thanks to (1) advancements in algorithm design and (2) big data and distributed computing. However, while I have taken every effort possible to ensure minimal errors within these risk calculations, they are not perfect. The measures were computed based on algorithms in a canned packed of R, and these methods leverage simulation.

In an attempt to strike a balance between efficiency of computing versus minimal errors, I had chosen a middle ground between these two extremes. After many preliminary tests to ensure the consistency of these simulation algorithms, I fixed the iteration sizes of the canned procedures to 5000 iterations. In larger networks, this may be considered too small. However, while I do claim this to be a limitation, I do not argue this is an inherent "problem" of the study in itself. Even within large networks, I had observed that the errors of the simulations were only as high as 0.5%. They were very efficient and consistent. Hence, while I do admit this is a possible limitation, I do not consider this to be of a major concern given the pre-testing conducted on these simulation procedures to ensure small error sizes.

6.4 Managerial Implications

The core purpose of this research was to gain a better understanding of the possible (now confirmed) relationship between inherent network structure and risk and risk propagation. When managers make decisions, they often do so in the face of uncertainty. They are not aware of all the consequences of their actions and decisions. While it may be cost or even initially "risk" effective to contract with a particular supplier, a lack of understanding the inherent structural changes to the supply network could lead to the opposite results a practitioner had expected.

This research presents a conceptual framework of network structure that helps to avoid the over complexity and vast number of possible measures of "a network". Such a conceptual framework will allow a manager to better understand the possible implications of their decisions from a structure-based view. In addition, they would be able to understand how might a change to their network (if they can characterize such a

change), impact their risk propagation levels. In addition, the framework presented in this study gives practitioners a collection of additional "dials" to turn, and to consider when attempting to optimize risk-mitigation policy.

The results of this study have illuminated just how turning those different dials will manifest into risk propagation. While some of those traditional dials such as "connectedness" may be appealing to turn and leverage, due to conventional wisdom at least, this study has established that these "dials" have their limits. There is no free lunch, as practitioners are well aware. When it comes to risk mitigation strategy, this study is the first to provide them a set of conceptual tools to use to consider within their decision making with respect to network structure.

Now that I have uncovered the static dynamics between structure and risk propagation, managers can attempt to use these to better understand which policies work best for them. In addition, this research has uncovered a dynamic between high and low dimensions of the supply network. At it's core, these dimensions are nothing more than subgraphs of the network. However, the light that I have cast on them illustrates that these dimensions have special purpose and meaning. When a firm makes a decision to change it's structure, this in effect will change the structure of the supply network, the ego-network, and the k-ego-network.

As demonstrated here, there are risk dynamics between these ego-networks. A strategy that may work well for an individual firm in regards to their risk may not work as well for the group (ego or more generally k-ego) as a whole. The question then for practitioners is how to best strike a balance between these paradoxical dynamics.

6.5 Future Research and Open Research Questions

As mentioned, this research is the first of it's kind to thoroughly understand the dynamics between structure and risk propagation. Throughout this dissertation, I have developed a few new tools of which I have leveraged to justify the individual parts of the primary inquiry of this thesis. As such, these smaller developments, along with the primary results, naturally lend themselves to unexplored boundaries that leave open

more questions than answers.

In regards to risk propagation and k-ego-networks, it would be worth while to further explore the competing dynamics of the effect that these structures have on risk at lower dimensions. Understanding the strategic implications would be of the utmost importance in this relatively new area of research, and I foresee many practical applications to determining optimal risk mitigation strategies.

In addition, it would be worth while to explore additional dynamics of the k-ego-networks, and, to explore the risk at these higher level networks. While it would be practically useless, what would be interesting to understand is just how far up the k-ego-network ladder one must consider before there is a diminishing return in investment of visibility and information seeking. Furthermore, understanding the dynamics of risk at these higher levels of analysis would enable us to better understand how risk "flows" from top to bottom, and under what conditions will the flow of this risk be thwarted.

Furthermore, throughout this research I leveraged a new construct that I designed known as the "people supply network". Ultimately, supply networks comprise of people. While we can create conceptualizations of abstract entities like "the firm" or "the supply network", at the heart of these constructs lie people, no different than how the atom lies at the heart of people. With the advent of social media and firms exceedingly moving towards social media to reach their market, understanding the dynamic interactions between people within the firm and those outside would allow for a clearer understanding of certain dynamics within supply networks that are still left to be explored. I have laid down a preliminary framework for this level of analysis. It may be worth while to develop a full and well argued framework for this level of analysis. Once designed, it's application to better understanding risk propagation within supply network's would easily be adopted. At the heart of the supply network and the various risks within that network lie the people.

Hence, I have motivated the following future questions that currently are left open to solve:

1. What are the dynamics between the varying dimensions of supply network structure and k-ego-network risk?

2. How does risk propagation evolve over time as the network evolves?

3. Do higher-order k-ego-networks experience higher levels of risk, and, if so, can collaborative strategies to mitigate this risk be beneficial, or harmful to the firm?

4. Under what conditions, given that network structure affects risk, would a firm find it beneficial to act opportunistically from a structural perspective by "tuning" the aforementioned dials?

5. What other properties of k-ego-networks affect other supply chain constructs such as power/trust/innovation, etc?

Chapter 7

Conclusion

Within this dissertation, I have sought to ascertain an understanding of a relationship between canonical network structure and risk. I feel that this dissertation has laid the theoretical and mathematical foundations to do just that. Many firm currently operate under the assumption that risk flows within a well defined collection of paths. I have argued that such a perspective is myopic, and doing so may lead to sub-optimal, if not non-optimal, decision making. Managerial decision making rests upon proper information available to said decision maker. If such information is inherently flawed, then it is logical to assume that the subsequent consequences of the decisions made will themselves lead to sub-optimal, or non-optimal, decisions.

As far as this author is concerned, no other researcher has taken upon the task to properly identify the core conceptual structural dimensional of a network (any network for that matter). This, as far as I can see, is very problematic. When a decision maker implements a well-defined construct that was designed along traditional strategic and managerial dimensions, what seems to be evident is that many are ignorant to the subsequent effects within the overall network that manifestations of such decision eventually come to form as a reality. If managers and practitioners alike took a more careful approach to their overall local, ego-network, and 2-ego-network designs, they may indeed achieve optimality in decision making. As I have demonstrated, decisions made within firms, from a network-structure perspective, affect the consequences of risk propagation within firms across the network, as well as within these tightly-formed groups, often coined "cliques".

As primary inquiry of which may be of the interest to the practitioner may indeed be "at what point shall my local benefit outweigh the group benefit?". I have demonstrated

that decisions pertaining to structural relationships between entities within a network do indeed play a determination towards that group's overall risk. At the end of the day, however, it is the individual firm that must weigh these benefits and risks, and make a determination as to what structural strategy such a firm must take in order to fulfill it's strategic objectives.

This dissertation has laid down both the mathematical and conceptual frameworks that will allow decision makers the tool-kit to be able to properly assess risk. Not only will they be able to assess risk from an internal perspective, but also of those more general networks (such as ego-network of 2-ego network, etc). Such an understanding of these dynamics will indeed leave the manager to consider not only what may be worthwhile to explore from a pure profit perspective, but also, from a risk perspective, of which portends potential issues within supply networks themselves. As such, this dissertation sets forth the primary conceptual framework for managers to leverage for the underlying objective of understanding the connection between their network structure and risk. If they happen to understand the dynamics between, lets say, centrality and risk, at different levels of the supply network, they may indeed pursue very different forms of strategic decision making than otherwise thought. Hence, this dissertation lends itself towards the advancement of strategic decision making.

In addition, this dissertation has implied that there exist multiple conceptualizations of "risk propagation". Namely, from the perspective of "causal", "diagnostic", and "change". This is the first of it's kind characterization of supply network risk propagation. Such a characterization could aid firms in deciding if they should behave more towards a group-based dynamic optimized solution or towards a firm-based optimized solution. In fact, I foresee future research in this field attempting to understanding the game-theoretic properties with respect to risk. Such a contribution would indeed be worthwhile, and would offer a strategic tool for firms that seek to optimize internal-based performance supply chain metrics.

In conclusion, I submit that this dissertation has set the theoretical conceptual foundation for much research to be explored. "Risk" has been throughly examined. However, *risk propagation* has not. This leads itself towards open questions, that have

yet to be explored. For example, could the various structural dynamics portend risk within networks, along varying degrees of analysis? Could there be a "social optimality where all parties, within a specified k-ego network benefit? And further, what would that k be?.

Appendix A

R-Code for Simulation

```
##################################################
#EC2 Functions
##################################################
library(aws.s3)
library(igraph)
library(bnlearn)
library(doParallel)
library(foreach)
library(brotli)
instanceID = 1
#Removed in publication for security reasons
Sys.setenv("AWS_ACCESS_KEY_ID" = "",
           "AWS_SECRET_ACCESS_KEY" = "",
           "AWS_DEFAULT_REGION" = "us-east-1")

ec2Work = function(instanceID){

  readBucket = paste("mgarvey",instanceID,sep="")
  writeBucket = switch(instanceID
    ,11,11,11,11,12,12,12,13,13,13,14,15,16,17,17,17)
  writeBucket = paste("mgarvey",writeBucket,sep="")
  if(instanceID==4)
    writeBucket = c(writeBucket,12)
  if(instanceID==7)
    writeBucket = c(writeBucket,13)

  toWork = function(){}
  if(instanceID<=10){
    toWork = ec2GenerateNetwork
```

```
29   }
30   else if(instanceID<=13){
31     toWork = ec2CalculateELR
32   }
33   else{
34     toWork = ec2CalculateMeasures
35   }
36
37   toWork(instanceID,readBucket,writeBucket)
38 }
39
40 ec2GenerateNetwork = function(instanceID,readBucket,writeBucket){
41   cores = detectCores()
42   cores = floor(cores)
43
44   cluster= makeCluster(cores)
45   registerDoParallel(cluster)
46
47   toGenerate = seq(max(100*(instanceID-1),100),100*instanceID,1)
48   toGen = c()
49   for(alpha in 10:1){
50     toGen = c(toGen,paste(toGenerate,"_",alpha,sep=""))
51   }
52
53   ret = foreach(i = toGen,.packages = c("bnlearn","igraph","brotli","
       aws.s3"),.export=ls(globalenv())) %dopar% {
54     splitNum = unlist(strsplit(i,"_"))
55     size = as.numeric(splitNum[1])
56     alpha = as.numeric(splitNum[2])
57
58     toret = list()
59     count = 1
60     for(beta in 1:10){
61       g=generateSupplyNetwork(size,floor(.7*size))
62       while(components(g)$no!=1)
63         g=generateSupplyNetwork(size,floor(.7*size))
64
```

```
      toret[[count]] = generateRiskGraph(g,alpha,beta)
      count=count+1
    }
    x=brotli_compress(serialize(toret,NULL))
    saveRDS(x,i)
    #s3saveRDS(x,i,ifelse(length(writeBucket)>1,sample(writeBucket,1)
    ,writeBucket))
    ""
  }
  stopCluster(cluster)
}

ec2CalculateELR = function(instanceID,readBucket,writeBucket){
  bList = list.files()
  countWait = 1
  cores = detectCores()
  cores = floor(.75*cores)
  #for(k in bList){
  if(TRUE){
    cluster= makeCluster(cores)
    registerDoParallel(cluster)

    ret = foreach(i = bList,.packages = c("bnlearn","igraph","brotli"
    ,"aws.s3"),.export=ls(globalenv())) %dopar% {
      x=readRDS(i)
      x=brotli_decompress(x)
      x=unserialize(x,NULL)
      toReturn = list()
      count = 1
      for(j in 1:length(x)){
        graphData = x[[j]]
        elrAll = calculateELRAll(graphData)
        riskMeasures = convertRiskMeasures(elrAll)
        toReturn[[count]] = c(graphData,riskMeasures)
        count = count+1
      }

```

```
100       x = serialize(toReturn,NULL)
101       x = brotli_compress(x)
102       saveRDS(x,paste(i,"_e",sep=""))
103       #s3saveRDS(x,i,writeBucket)
104       ""
105     }
106     stopCluster(cluster)
107     #bList = get_bucket_df(readBucket)$Key
108   }
109 }
110
111 ec2CalculateMeasures = function(instanceID,readBucket,writeBucket){
112   dumpBucket="mgarveydumpbucket"
113   bList = get_bucket_df(readBucket)$Key
114   countWait = 1
115   cores = detectCores()
116   cores = floor(.75*cores)
117   while(TRUE){
118     while((length(bList)<cores&&countWait<3)){
119       Sys.sleep(30)
120       bList = get_bucket_df(readBucket)$Key
121       countWait = countWait+1
122     }
123
124     if(length(bList)==0){
125       next
126       countWait = 1
127     }
128     countWait=1
129     cluster= makeCluster(cores)
130     registerDoParallel(cluster)
131
132     ret = foreach(i = bList,.packages = c("bnlearn","igraph","brotli"
      ,"aws.s3"),.export=ls(globalenv())) %dopar% {
133       x=s3readRDS(i,readBucket)
134       x=brotli_decompress(x)
135       x=unserialize(x,NULL)
```

```
136
137       graphData = x[[1]]
138       elrAll = getAllMetrics(graphData)
139       e1 = elrAll[["egodata1"]]
140       e2 = elrAll[["egodata2"]]
141       loc = elrAll[["localData"]]
142
143       for(j in 2:length(x)){
144         graphData = x[[j]]
145         elrAll = getAllMetrics(graphData)
146         e1 = rbind(e1,elrAll[["egodata1"]])
147         e2 = rbind(e2,elrAll[["egodata2"]])
148         loc = rbind(loc,elrAll[["localData"]])
149       }
150
151       x = serialize(e1,NULL)
152       x = brotli_compress(x)
153       s3saveRDS(x,paste("ego1_",i,sep=""),writeBucket)
154
155       x = serialize(e2,NULL)
156       x = brotli_compress(x)
157       s3saveRDS(x,paste("ego2_",i,sep=""),writeBucket)
158
159       x = serialize(loc,NULL)
160       x = brotli_compress(x)
161       s3saveRDS(x,paste("loc_",i,sep=""),writeBucket)
162
163       delete_object(i,readBucket)
164       put_object(i,dumpBucket)
165       ""
166     }
167     stopCluster(cluster)
168     bList = get_bucket_df(readBucket)$Key
169   }
170
171 }
172
```

```
ec2CombineMeasures = function(instanceID){

}




##################################################
#Utility Functions
##################################################
getDesc=function(graph,node){
  #First, obtain the adjacency matrix
  adj_matrix = amat(graph)

  #Next, create a list of names
  node_labels = colnames(adj_matrix)

  #Create and initialize list of nodes to further explore
  to_explore = colnames(adj_matrix)[which(adj_matrix[node,]==1)]

  #Now, create an empty vector that will store the decendent names
  desc = c()

  #Create list of explored
  explored = c(node)

  #if there are no children, just return an empty list
  if(length(to_explore)==0)
    return(c())

  #While there are nodes left to explore, account for children
  while(length(to_explore)>0){
    #Set the next node to explore, add it to the desc list and
   explored list
    next_node = to_explore[1]
    desc=c(desc,next_node)
```

```
    explored=c(explored,next_node)

    #Now, obtain the children for this node
    children = colnames(adj_matrix)[which(adj_matrix[next_node,]==1)]

    #Iterate through the children.  If we didn't explore one, add it
   to the list
    for(i in children){
      if(!is.element(i,explored)){
        to_explore=c(to_explore,i)
      }
    }

    #Last, remove the first entry since we just explored it
    to_explore=to_explore[-1]
  }
  desc=intersect(desc,desc)
  return(desc)
}
getAllDesc = function(graph){
  to_return = list()
  to_return[["graph"]] = graph

  nodes = colnames(amat(graph))

  for(i in nodes){
    desc = getDesc(graph,i)

    to_return[[i]] = desc
  }

  return(to_return)

}
getAnces = function(graph,node){
  #First, obtain the adjacency matrix
  adj_matrix = amat(graph)
```

```
#Next, create a list of names
node_labels = colnames(adj_matrix)

#Create and initialize list of nodes to further explore
to_explore = colnames(adj_matrix)[which(adj_matrix[,node]==1)]

#Now, create an empty vector that will store the ancestor names
ances = c()

#Create list of explored
explored = c(node)

#if there are no parents, just return an empty list
if(length(to_explore)==0)
  return(c())

#While there are nodes left to explore, account for children
while(length(to_explore)>0){
  #Set the next node to explore, add it to the desc list and
  explored list
  next_node = to_explore[1]
  ances=c(ances,next_node)
  explored=c(explored,next_node)

  #Now, obtain the children for this node
  parents = colnames(adj_matrix)[which(adj_matrix[,next_node]==1)]

  #Iterate through the children.  If we didn't explore one, add it
  to the list
  for(i in parents){
    if(!is.element(i,explored)){
      to_explore=c(to_explore,i)
    }
  }

  #Last, remove the first entry since we just explored it
```

```
    to_explore=to_explore[-1]
  }
  ances=intersect(ances,ances)
  return(ances)
}
getAllAnces = function(graph){
  to_return = list()
  to_return[["graph"]] = graph
  nodes = colnames(amat(graph))
  for(i in nodes){
    ances = getAnces(graph,i)
    to_return[[i]] = ances
  }
  return(to_return)
}
getRoots = function(graph,isMatrix=FALSE){
  m=graph

  if(!isMatrix)
    m = as_adjacency_matrix(graph,sparse=FALSE)
  if(length(m)<=1)
    return(m)
  roots = which(apply(m,2,sum)==0)
  return(roots)
}
getParents = function(graph,node){
  return(graph$nodes[[node]]$parents)
}
getChildren = function(graph,node){
  return(graph$nodes[[node]]$children)
}

getAllCycles = function(graph){
  adj = as_adj(graph,sparse=FALSE)

  nodes = as.vector(V(graph))
  for(i in as.vector(V(graph))){
```

```
    dfsr = as.vector(dfs(graph,i)$order)
    adjTemp = matrix(rep(0,length(V(graph))),nrow=length(V(graph)))
    for(j in 1:(length(dfsr))-1){
      if(adj[dfsr[j],dfsr[j+1]]==1)
        adjTemp[dfsr[j],dfsr[j+1]]=1
    }

    toExtract = adj - adjTemp
    toExtract[upper.tri(toExtract)] = 0
    missingEdges = which(toExtract!=0)

    for(k in missingEdge){
      row = k%%length(nodes)
      column = ceiling(k/length(nodes))
      edge = paste(row,"-",column)

    }
  }

}
kEgoConnectivity = function(graph,node){
  otherNodes = setdiff(names(V(graph)),node)
  k = components(induced_subgraph(graph,otherNodes))$no
  level = 0
  while(k==1&&length(otherNodes)>0){
    level=level+1
    ego = getKEgo(graph,level,node)
    otherNodes = setdiff(names(V(graph)),names(V(ego)))
    k = components(induced_subgraph(graph,otherNodes))$no
  }
  return(level)
}
countTriads = function(graph,node){
  #First, enumerate the neighbors
  neighbors = names(neighbors(graph,node))
  if(length(neighbors)<2)
    return(list(count=0,pairs=list()))
```

```
  adjM = as_adj(graph,sparse=FALSE)
  pairs = list()
  count = 0
  for(i in 1:(length(neighbors)-1)){
    for(j in i:length(neighbors)){
      p1 = neighbors[i]
      p2 = neighbors[j]
      if(adjM[p1,p2]==1){
        count = count+1
        pair = c(p1,p2)
        pairs[[count]] = pair
      }
    }
  }
  return(list(count=count,pairs=pairs))
}
convertGraph = function(graph){
  if(class(graph)=="igraph"){
    toRet = empty.graph(paste("n",1:length(V(graph)),sep=""))
    amat(toRet) = as_adj(graph)
  }
  return(graph_from_adjacency_matrix(amat(graph)))
}

findStructuralHoles = function(graph,node){
  #First, enumerate the neighbors
  neighbors = names(neighbors(graph,node))
  if(length(neighbors)<2)
    return(list(count=0,pairs=list()))
  adjM = as_adj(graph,sparse=FALSE)
  count = 0
  triads = list()
  for(i in 1:(length(neighbors)-1)){
    for(j in (i+1):length(neighbors)){
      p1 = neighbors[i]
      p2 = neighbors[j]
      if(adjM[p1,p2]==0){
```

```
        count = count+1
        pair = c(p1,p2)
        triads[[paste(node,"-",p1,"-",p2,sep="")]]=list(missing=c(p1,
    p2))
      }
    }
  }
  return(list(count=count,triads=triads))
}
findAllStructuralHoles = function(graph){
  sh=list()
  nodes=V(graph)$name
  for(i in nodes){
    sh=c(sh,findStructuralHoles(graph,i)$triads)
  }
  return(sh)
}
splitVector = function(x,m){
  ret = list()
  count = 1
  gSize = floor(length(x)/m)
  for(i in 1:m){
    start = (i-1)*gSize+1
    end = i*gSize
    ret[[count]] = x[start:end]
    count = count+1
  }
  rem = length(x)%%m
  index = length(x)
  if(rem==0)
    return(ret)

  for(i in 1:rem){
    ni = sample(1:length(ret),1)
    ret[[ni]] = c(ret[[ni]],x[index])
    index = index-1
  }
```

```
  index = 1
  for(i in 1:length(ret)){
    for(j in 1:length(ret[[i]])){
      ret[[i]][j] = index
      index = index+1
    }
  }
  return(ret)
}
findLayer = function(layers,node){
  for(i in 1:length(layers)){
    if(is.element(node,layers[[i]]))
      return(i)
  }
}
findOnes = function(matrix){
  x = which(matrix==1)
  f = x[1]
  x=x[-1]
  size = nrow(matrix)
  row = f%%size
  row = ifelse(row==0,3,row)
  col = ceiling(f/size)
  toRet = c(row,col)

  for(i in x){
    f = i
    row = f%%size
    row = ifelse(row==0,3,row)
    col = ceiling(f/size)
    toRet = rbind(toRet,c(row,col))
  }
  colnames(toRet)=c("from","to")
  return(toRet)
}

getEdgesFromGraph = function(graph,nodes){
```

```
}
##################################################
##################################################

##################################################
#Primary Calculation Functions
##################################################
calculateProb=function(graph,fit,node){
  cmd = paste("cpquery(fit,event=(",node,"==1),evidence=TRUE)",sep="")
  prob = eval(parse(text=cmd))
}
calculateER = function(graph,fit,node,value){
  ev = paste("(",node,"==",value,")",sep="")
  des = getDesc(graph,node)
  anc = getAnces(graph,node)
  nodes=union(des,anc)
  cmd = paste("cpdist(fit,nodes,evidence=",ev,")",sep="")
  samps = apply(eval(parse(text=cmd)),2,as.numeric)

  #getDescedents
  dsamps = apply(t(t(samps[,des])),1,sum)

  #getAcen
  asamps = apply(t(t(samps[,anc])),1,sum)

  samps = matrix(c(dsamps,asamps),ncol=2)

  vels = apply(samps,1,function(x){
    if(x[2]==0)
      return(x[1])
    else{
      return((x[1]-x[2])/x[2])
    }
  })

```

```
  ercf = mean(dsamps)
  ersf = mean(asamps)
  erv = mean(vels)

  return(list(ercf=ercf,ersf=ersf,erv=erv))
}
calculateELR = function(graph,fit,node){
  prob = calculateProb(graph,fit,node)
  r0 = calculateER(graph,fit,node,0)
  r1 = calculateER(graph,fit,node,1)

  elrcf = (1+r1[["ercf"]])*prob + r0[["ercf"]]*(1-prob)
  elrsf = (1+r1[["ersf"]])*prob + r0[["ersf"]]*(1-prob)
  elrv  = r1[["erv"]]*prob + r0[["erv"]]*(1-prob)


  return(list(elrcf = elrcf,elrsf=elrsf,elrv = elrv))
}
calculateELRAll = function(graphData,multi=FALSE){
  graph = graphData[["graph"]]
  fit = graphData[["fit"]]
  if(multi){
    nodes = names(graph$nodes)
    cl = makeCluster(detectCores()/2)
    registerDoParallel(cl)
    ret=list()
    ret = foreach(i = nodes, .final = function(x) setNames(x, nodes)
    ,.packages = "bnlearn",.export=ls(globalenv())) %dopar% {
      calculateELR(graph,fit,i)
    }
    stopCluster(cl)
    stopImplicitCluster()
    return(ret)
  }
  nodes = names(graph$nodes)
  ret=list()
```

```
  for(i in nodes){
    ret[[i]]=calculateELR(graph,fit,i)
  }
  return(ret)
}
getAllMetrics = function(graphData){
  #First, get 2-ego level metrics:
  sNetwork = graphData$supplyNetwork
  nodes = c(V(sNetwork)$name,E(sNetwork)$name)
  egodata2 = getEgoMetrics(graphData,2,nodes[1])
  node = nodes[1]
  nodes=nodes[-1]
  gd = graphData
  gd[["2egoRisk"]] = list()
  gd[["2egoRisk"]][[node]]=list(avg_elrcf=egodata2[,"avg_elrcf"],avg_
    elrsf=egodata2[,"avg_elrsf"],avg_elrv=egodata2[,"avg_elrv"],sd_
    elrcf=egodata2[,"sd_elrcf"],sd_elrsf=egodata2[,"sd_elrsf"],sd_
    elrv=egodata2[,"sd_elrv"],cv_elrcf=egodata2[,"cv_elrcf"],cv_elrsf
    =egodata2[,"cv_elrsf"],cv_elrv=egodata2[,"cv_elrv"])
  count = 1

  for(i in nodes){
    egodata2=rbind(egodata2,getEgoMetrics(gd,2,i))
    count=count+1
    gd[["2egoRisk"]][[i]]=list(avg_elrcf=egodata2[count,"avg_elrcf"],
    avg_elrsf=egodata2[count,"avg_elrsf"],avg_elrv=egodata2[count,"
    avg_elrv"],sd_elrcf=egodata2[count,"sd_elrcf"],sd_elrsf=egodata2[
    count,"sd_elrsf"],sd_elrv=egodata2[count,"sd_elrv"],cv_elrcf=
    egodata2[count,"cv_elrcf"],cv_elrsf=egodata2[count,"cv_elrsf"],cv
    _elrv=egodata2[count,"cv_elrv"])
  }
  count = 1
  egodata = getEgoMetrics(gd,1,node)
  gd[["egoRisk"]] = list()
```

```
  gd[["egoRisk"]][[node]]=list(avg_elrcf=egodata[,"avg_elrcf"],avg_elrsf=egodata[,"avg_elrsf"],avg_elrv=egodata[,"avg_elrv"],sd_elrcf=egodata[,"sd_elrcf"],sd_elrsf=egodata[,"sd_elrsf"],sd_elrv=egodata[,"sd_elrv"],cv_elrcf=egodata[,"cv_elrcf"],cv_elrsf=egodata[,"cv_elrsf"],cv_elrv=egodata[,"cv_elrv"])


  for(i in nodes){
    egodata=rbind(egodata,getEgoMetrics(gd,1,i))
    count=count+1
    gd[["egoRisk"]][[i]] =
      list(avg_elrcf=egodata[count,"avg_elrcf"],avg_elrsf=egodata[count,"avg_elrsf"],avg_elrv=egodata[count,"avg_elrv"],sd_elrcf=egodata[count,"sd_elrcf"],sd_elrsf=egodata[count,"sd_elrsf"],sd_elrv=egodata[count,"sd_elrv"],cv_elrcf=egodata[count,"cv_elrcf"],cv_elrsf=egodata[count,"cv_elrsf"],cv_elrv=egodata[count,"cv_elrv"])

  }

  localData = getLocalMetrics(gd,node)
  for(i in nodes){
    localData = rbind(localData,getLocalMetrics(gd,i))
  }

  return(list(egodata2=egodata2,egodata1=egodata,localData=localData))
}


getLocalMetrics = function(graphData,node){
  graph = graphData[["supplyNetwork"]]
  egoRisk = graphData[["egoRisk"]]
  if(substr(node,1,1)=="e")
    graph = graphData[["lineGraph"]]


```

```
nBetweenness = betweenness(graph,node)
nCloseness = closeness(graph,node)
nEigen = eigen_centrality(graph,node)$vector[node]
nDegree = degree(graph,node)
nEccentricity = eccentricity(graph,node)
#First, find the center
me = eccentricity(graph)
graphCenter = names(which(me == min(me)))
nMinCenterDistance = min(distances(graph,node,graphCenter))
nMaxCenterDistance = max(distances(graph,node,graphCenter))

nDist = distances(graph,node)
nAvgSPathDis = sum(nDist)/(length(nDist)-1)

nDensity = nDegree/vcount(graph)
nClusterCoeff = transitivity(graph,"local",node,isolates="zero")
nCoreness = coreness(graph)[node]
nTriads = countTriads(graph,node)$count
nSH = choose(nDegree,2)-nTriads
nEgoConnectivity = kEgoConnectivity(graph,node)
isNode = ifelse(substr(node,1,1)=="n",1,0)
elrsf = graphData[["elrsf"]][node]
elrcf = graphData[["elrcf"]][node]
elrv = graphData[["elrv"]][node]
#get av,sd and cv of elrcf, elrsf and elrrv for 2-ego
avg_elrcf_ego = egoRisk[[node]][["avg_elrcf"]]
avg_elrsf_ego = egoRisk[[node]][["avg_elrsf"]]
avg_elrv_ego = egoRisk[[node]][["avg_elrv"]]

sd_elrcf_ego = egoRisk[[node]][["sd_elrcf"]]
sd_elrsf_ego = egoRisk[[node]][["sd_elrsf"]]
sd_elrv_ego = egoRisk[[node]][["sd_elrv"]]

cv_elrcf_ego = egoRisk[[node]][["cv_elrcf"]]
cv_elrsf_ego = egoRisk[[node]][["cv_elrsf"]]
cv_elrv_ego = egoRisk[[node]][["cv_elrv"]]

```

```
  alpha = graphData[["alpha"]]
  beta = graphData[["beta"]]

  toRet = cbind(alpha,beta,avg_elrcf_ego,avg_elrsf_ego,avg_elrv_ego,
    sd_elrcf_ego,sd_elrsf_ego,sd_elrv_ego,cv_elrcf_ego,cv_elrsf_ego,
    cv_elrv_ego,isNode,nBetweenness,nCloseness,nEigen,nDegree,
    nEccentricity,nMinCenterDistance,nMaxCenterDistance,nAvgSPathDis,
    nDensity,nClusterCoeff,nCoreness,nTriads,nSH,nEgoConnectivity,
    elrsf,elrcf,elrv)

  return(toRet)
}




getEgoMetrics = function(graphData,k,node){
  sNetwork = graphData[["supplyNetwork"]]
  graph = graphData[["graph"]]
  elrcf = graphData[["elrcf"]]
  elrsf = graphData[["elrsf"]]
  elrv = graphData[["elrv"]]
  riskMeasures = list()
  if(k==1){
    riskMeasures = graphData[["2egoRisk"]]
  }


  #Core Graph Metrics
  egoN = getKEgo(sNetwork,k,node)
  metrics = getEgoGraphMetrics(egoN)


  nodes = as_ids(V(egoN))
  nodes = c(nodes,edge_attr(egoN,"name",E(egoN)))
  #Now, find the ELRCF for the ego network
  elrcf_vals = elrcf[nodes]
```

```
avg_elrcf = mean(elrcf_vals)
sd_elrcf = sd(elrcf_vals)
cv_elrcf = sd_elrcf/avg_elrcf

#Ditto for the ELRSF
elrsf_vals = elrsf[nodes]
avg_elrsf = mean(elrsf_vals)
sd_elrsf = sd(elrsf_vals)
cv_elrsf = sd_elrsf/avg_elrsf

#Last, the velocity
elrv_vals = elrv[nodes]
avg_elrv = mean(elrv_vals)
sd_elrv = sd(elrv_vals)
cv_elrv = sd_elrv/avg_elrv

alpha = graphData[["alpha"]]
beta = graphData[["beta"]]
metrics = cbind(alpha,beta,metrics,avg_elrcf,avg_elrsf,avg_elrv,sd_
  elrcf,sd_elrsf,sd_elrv,cv_elrcf,cv_elrsf,cv_elrv)

if(k==1){
  #get av,sd and cv of elrcf, elrsf and elrrv for 2-ego
  avg_elrcf_2ego = riskMeasures[[node]][["avg_elrcf"]]
  avg_elrsf_2ego = riskMeasures[[node]][["avg_elrsf"]]
  avg_elrv_2ego = riskMeasures[[node]][["avg_elrv"]]

  sd_elrcf_2ego = riskMeasures[[node]][["sd_elrcf"]]
  sd_elrsf_2ego = riskMeasures[[node]][["sd_elrsf"]]
  sd_elrv_2ego = riskMeasures[[node]][["sd_elrv"]]

  cv_elrcf_2ego = riskMeasures[[node]][["cv_elrcf"]]
  cv_elrsf_2ego = riskMeasures[[node]][["cv_elrsf"]]
  cv_elrv_2ego = riskMeasures[[node]][["cv_elrv"]]

```

```
    metrics=cbind(avg_elrcf_2ego,avg_elrsf_2ego,avg_elrv_2ego,sd_elrcf_2ego,sd_elrsf_2ego,sd_elrv_2ego,cv_elrcf_2ego,cv_elrsf_2ego,cv_elrv_2ego,metrics)
    return(metrics)
  }
  return(metrics)
}

#More of a utility for getEgoMetrics
getEgoGraphMetrics = function(egoNetwork){
  egoN = egoNetwork
  eNNodes = length(V(egoN))
  eNArcs = length(E(egoN))
  eEdgeConnectivity = edge_connectivity(egoN)
  eNodeConnectivity = vertex_connectivity(egoN)
  eDiameter = diameter(egoN)
  eRadius = radius(egoN)
  eGirth = girth(egoN)$girth
  eCliqueNum = clique_num(egoN)
  eNumOfLargestCliques = length(largest_cliques(egoN))
  eAvCliqueSize = mean(unlist(lapply(cliques(egoN),length)))
  eNSH = triad_census(egoN)[11]
  eNTriads = triad.census(egoN)[16]
  eAvgConstraint = mean(constraint(egoN))
  eConstraintCen = centralize(constraint(egoN),normalized=FALSE)
  eDensity = edge_density(egoN)
  eClusterCoeff = transitivity(egoN,isolates="zero")
  eClusterCoeffAv = transitivity(egoN,type="average",isolates="zero")
  eAvgShortestPathDis = mean_distance(egoN)
  eAvgBetweeness = mean(betweenness(egoN))
  eAvgCloseness = mean(closeness(egoN))
  eAvgDegree = mean(degree(egoN))
  eAvgEV = mean(eigen_centrality(egoN)$vector)
  eCenBetweeness = centr_betw(egoN,normalized=FALSE)$centralization
  eCenEigen = centr_eigen(egoN,normalized=FALSE)$centralization
  eCenCloseness = centr_clo(egoN,normalized=FALSE)$centralization
  eCenDegree = centr_degree(egoN,normalized=FALSE)$centralization
```

```
  eAvgEccentricity = mean(eccentricity(egoN))
  eEccentricityCen = centralize(eccentricity(egoN),normalized=FALSE)
  ret=cbind(eNNodes,eNArcs,eEdgeConnectivity,eNodeConnectivity,
    eDiameter,eRadius,eGirth,eCliqueNum,eNumOfLargestCliques,
    eAvCliqueSize,eNSH,eNTriads,eAvgConstraint,eConstraintCen,
    eDensity,eClusterCoeff,eClusterCoeffAv,eAvgShortestPathDis,
    eAvgBetweeness,eAvgCloseness,eAvgDegree,eAvgEV,eCenBetweeness,
    eCenEigen,eCenCloseness,eCenDegree,eAvgEccentricity,
    eEccentricityCen)
  return(ret)
}

convertRiskMeasures = function(measureData){
  theNodes = names(measureData)
  newList = list()
  elrcf = c()
  elrsf = c()
  elrv = c()

  for(i in theNodes){
    measures = measureData[[i]]
    elrcf=c(elrcf,measures$elrcf)
    elrsf = c(elrsf,measures$elrsf)
    elrv = c(elrv, measures$elrv)
  }

  names(elrcf) = theNodes
  names(elrsf) = theNodes
  names(elrv) = theNodes

  return(list(elrcf=elrcf,elrsf=elrsf,elrv=elrv))
}

##################################################
#Network Generation Functions
##################################################
getKEgo=function(graph,k,node){
```

```
  sNetwork = graph

  if(substr(node,1,1)=="n"){
    nodes = c(substr(node,2,nchar(node)))
    nodes = as.numeric(nodes)
    for(i in 1:k){
      nodesTemp = nodes
      for(j in nodes){
        nodesTemp = union(nodesTemp,as.vector(neighbors(sNetwork,j,"
    all")))
      }
      nodes = nodesTemp
    }
    return(induced_subgraph(sNetwork,nodes))
  }else{
    edgeNum = as.numeric(substr(node,2,nchar(node)))
    vert = get.edges(sNetwork,E(sNetwork))[edgeNum,]
    vert = paste("n",vert,sep="")
    g1 = getKEgo(graph,k,vert[1])
    g2 = getKEgo(graph,k,vert[2])
    tNodes = union(as_ids(V(g1)),as_ids(V(g2)))
    return(induced_subgraph(sNetwork,tNodes))
  }
}

populateUniform=function(graph){
  nodes = names(graph$nodes)
  dis = list()
  for(i in nodes){
    print(i)
    parents = getParents(graph,i)

    numC = 2^(length(parents)+1)
    x=rep(.5,numC)
    dim(x) = rep(2,length(parents)+1)
    x_lab = list()
    x_lab[[i]] = c("TRUE","FALSE")
```

```
    for(j in parents){
      x_lab[[j]] = c("TRUE","FALSE")
    }
    dimnames(x) = x_lab
    dis[[i]] = x
  }

  dfit = custom.fit(graph,dis)
  return(dfit)
}
populateUniformDiff=function(graph,prob){
  nodes = names(graph$nodes)
  dis = list()
  for(i in nodes){
    print(i)
    parents = getParents(graph,i)

    numC = 2^(length(parents)+1)
    x=rep(c(prob,1-prob),numC/2)
    dim(x) = rep(2,length(parents)+1)
    x_lab = list()
    x_lab[[i]] = c("TRUE","FALSE")
    for(j in parents){
      x_lab[[j]] = c("TRUE","FALSE")
    }
    dimnames(x) = x_lab
    dis[[i]] = x
  }

  dfit = custom.fit(graph,dis)
  return(dfit)
}
populateBeta=function(graph,a,b){
  nodes = names(graph$nodes)
  dis = list()

  for(i in nodes){
```

```
    print(i)
    parents = getParents(graph,i)

    numC = 2^(length(parents)+1)
    numToGen = numC/2

    x=rep(0,numC)
    probs = round(rbeta(numToGen,a,b),2)*100
    odd = which(probs%%2!=0)
    probs[odd]=probs[odd]+1
    probs=probs/100
    antiProbs = 1-probs
    x[seq(1,numC,2)] = probs
    x[seq(2,numC,2)] = antiProbs
    dim(x) = rep(2,length(parents)+1)
    x_lab = list()
    x_lab[[i]] = c(1,0)
    for(j in parents){
      x_lab[[j]] = c(1,0)
    }
    dimnames(x) = x_lab
    dis[[i]] = x
  }

  dfit = custom.fit(graph,dis)
  return(dfit)

}
populateSame=function(graph,prob){
  nodes = names(graph$nodes)
  dis = list()

  for(i in nodes){
    print(i)
    parents = getParents(graph,i)

    numC = 2^(length(parents)+1)
```

```
    numToGen = numC/2

    x=rep(0,numC)
    probs = rep(prob,numToGen)
    antiProbs = 1-probs
    x[seq(1,numC,2)] = probs
    x[seq(2,numC,2)] = antiProbs
    dim(x) = rep(2,length(parents)+1)
    x_lab = list()
    x_lab[[i]] = c(1,0)
    for(j in parents){
      x_lab[[j]] = c(1,0)
    }
    dimnames(x) = x_lab
    dis[[i]] = x
  }

  dfit = custom.fit(graph,dis)
  return(dfit)

}

generateSupplyNetwork = function(n,horizontal = 5){
  vertical = 5
  nodesPerLevel = floor(n/horizontal)

  graph=make_empty_graph()
  graph = add_vertices(graph,n,attr=list(name=paste("n",1:n,sep="")))
  layers = splitVector(1:n,horizontal)
  for(i in 2:horizontal){
    nodes = layers[[i]]
    potentialParents = 1:(min(nodes)-1)
    for(j in nodes){
      dist = abs(rnorm(10,.5,.25))
      dist = dist/sum(dist)
      pSize = sample(1:min(10,length(potentialParents),1,prob = dist)
    )
```

```
      nP = sample(1:length(potentialParents),min(pSize,length(potentialParents)))
      for(k in 1:pSize){
        graph = add_edges(graph,c(nP[k],j))
      }
    }
  }
  eNames = c()
  edges = get.edges(graph,E(graph))
  for(i in 1:nrow(edges)){
    eNames = c(eNames,paste("e",i,sep=""))
  }
  graph = set_edge_attr(graph,"name",E(graph),eNames)

  sh1=findAllStructuralHoles(graph)
  ta1 = sample(names(sh1),sample(5:length(names(sh1))))
  visited = c()
  for(i in ta1){
    nums = sh1[[i]]$missing
    if(is.element(paste(nums,collapse=""),visited))
      next
    num=sapply(nums,function(x){as.numeric(substr(x,2,nchar(x)))})
    nEdge = length(E(graph))+1
    if(num[1]<num[2]){
      graph=add_edges(graph,num,attr=list(name=paste("e",nEdge,sep=""))))
      nEdge = nEdge+1
      visited = c(visited,paste(nums,collapse=""),paste(rev(nums),collapse=""))
    }
  }

  return(graph)
}
generateRiskGraph = function(supplyNetwork,alpha,beta){

  nNodes = length(V(supplyNetwork))
```

```
  nEdges = length(E(supplyNetwork))
  nodes = paste("n",1:nNodes,sep="")
  eNodes = paste("e",1:nEdges,sep="")

  nodes = c(nodes,eNodes)

  graph = empty.graph(nodes)
  mat = amat(graph)
  edges = get.edges(supplyNetwork,E(supplyNetwork))
  for(i in 1:nrow(edges)){
    e=edges[i,]
    v = paste("n",e,sep="")
    mat[v[1],v[2]] = 1
    mat[v[1],paste("e",i,sep="")] = 1
    mat[paste("e",i,sep=""),v[2]] = 1
  }
  amat(graph) = mat
  fit = populateBeta(graph,alpha,beta)
  supplyNetwork = as.undirected(supplyNetwork)
  lineGraph = line.graph(supplyNetwork)
  lineGraph = set_vertex_attr(lineGraph,"name",V(lineGraph),E(
    supplyNetwork)$name)

  return(list(graph = graph,supplyNetwork = supplyNetwork, lineGraph
    = lineGraph, alpha = alpha,beta = beta,fit = fit))
}

##################################################
#Demonstrations/Simulations
##################################################
demonstration = function(){
  nodes=paste("n",1:9,sep="")
  g1=empty.graph(nodes)
  g2=empty.graph(nodes)
  g3=empty.graph(nodes)
  g4=empty.graph(nodes)

```

```
m=matrix(rep(0,81),nrow=9)
m[1,2]=1
m[2,3]=1
m[3,4]=1
m[4,5]=1
m[5,6]=1
m[6,7]=1
m[7,8]=1
m[8,9]=1
amat(g1)=m
m=matrix(rep(0,81),nrow=9)
m[1,5]=1
m[2,5]=1
m[3,5]=1
m[4,5]=1
m[5,6]=1
m[5,7]=1
m[5,8]=1
m[5,9]=1
amat(g2)=m
m=matrix(rep(0,81),nrow=9)
m[1,2]=1
m[1,3]=1
m[2,4]=1
m[2,5]=1
m[4,8]=1
m[3,6]=1
m[3,7]=1
m[7,9]=1
amat(g3)=m
m=matrix(rep(0,81),nrow=9)
m[1,2]=1
m[1,3]=1
m[2,4]=1
m[4,6]=1
m[6,8]=1
m[3,5]=1
```

```
m[5,7]=1
m[7,9]=1
amat(g4)=m

gi1=graph_from_adjacency_matrix(amat(g1),mode="undirected")
gi2=graph_from_adjacency_matrix(amat(g2),mode="undirected")
gi3=graph_from_adjacency_matrix(amat(g3),mode="undirected")
gi4=graph_from_adjacency_matrix(amat(g4),mode="undirected")

results = matrix(rep(0,9),nrow=1,byrow=TRUE)
tNames = c("radius","diameter","abc","acc","aec","aelrcf","sdelrcf"
  ,"cvelrcf","prob")
fb1 = c(radius(gi1),diameter(gi1),mean(betweenness(gi1)),mean(
  closeness(gi1)),mean(eigen_centrality(gi1)$vector))
fb2 = c(radius(gi2),diameter(gi2),mean(betweenness(gi2)),mean(
  closeness(gi2)),mean(eigen_centrality(gi2)$vector))
fb3 = c(radius(gi3),diameter(gi3),mean(betweenness(gi3)),mean(
  closeness(gi3)),mean(eigen_centrality(gi3)$vector))
fb4 = c(radius(gi4),diameter(gi4),mean(betweenness(gi4)),mean(
  closeness(gi4)),mean(eigen_centrality(gi1)$vector))

for(i in seq(0.01,.99,.01)){
  if(100*i%%10==0)
    print(i)
  f1=populateSame(g1,i)
  f2=populateSame(g2,i)
  f3=populateSame(g3,i)
  f4=populateSame(g4,i)

  tc = list(f1,f2,f3,f4)

  r1 = convertRiskMeasures(calculateELRAll(list(graph=g1,fit=f1)))
  r2 = convertRiskMeasures(calculateELRAll(list(graph=g2,fit=f2)))
  r3 = convertRiskMeasures(calculateELRAll(list(graph=g3,fit=f3)))
  r4 = convertRiskMeasures(calculateELRAll(list(graph=g4,fit=f4)))

  e1 = unlist(r1$elrcf)
```

```
    e2 = unlist(r2$elrcf)
    e3 = unlist(r3$elrcf)
    e4 = unlist(r4$elrcf)

    m1 = mean(e1)
    m2 = mean(e2)
    m3 = mean(e3)
    m4 = mean(e4)

    s1 = sd(e1)
    s2 = sd(e2)
    s3 = sd(e3)
    s4 = sd(e4)

    cv1 = s1/m1
    cv2 = s2/m2
    cv3 = s3/m3
    cv4 = s4/m4
    ne1 = c(fb1,m1,s1,cv1,i)
    ne2 = c(fb2,m2,s2,cv2,i)
    ne3 = c(fb3,m3,s3,cv3,i)
    ne4 = c(fb4,m4,s4,cv4,i)

    results=rbind(results,ne1,ne2,ne3,ne4)
    results=results[-1,]
  }
  df=data.frame(results)
  names(df)=tNames
  return(df)
}
demonstration2 = function(){
  g1=empty.graph(nodes)
  g2=empty.graph(nodes)
  g3=empty.graph(nodes)
  g4=empty.graph(nodes)

  m=matrix(rep(0,81),nrow=9)
```

```
m[1,2]=1
m[2,3]=1
m[3,4]=1
m[4,5]=1
m[5,6]=1
m[6,7]=1
m[7,8]=1
m[8,9]=1
amat(g1)=m
m=matrix(rep(0,81),nrow=9)
m[1,5]=1
m[2,5]=1
m[3,5]=1
m[4,5]=1
m[5,6]=1
m[5,7]=1
m[5,8]=1
m[5,9]=1
amat(g2)=m
m=matrix(rep(0,81),nrow=9)
m[1,2]=1
m[1,3]=1
m[2,4]=1
m[2,5]=1
m[4,8]=1
m[3,6]=1
m[3,7]=1
m[7,9]=1
amat(g3)=m
m=matrix(rep(0,81),nrow=9)
m[1,2]=1
m[1,3]=1
m[2,4]=1
m[4,6]=1
m[6,8]=1
m[3,5]=1
m[5,7]=1
```

```
m[7,9]=1
amat(g4)=m

gi1=graph_from_adjacency_matrix(amat(g1),mode="undirected")
gi2=graph_from_adjacency_matrix(amat(g2),mode="undirected")
gi3=graph_from_adjacency_matrix(amat(g3),mode="undirected")
gi4=graph_from_adjacency_matrix(amat(g4),mode="undirected")
detach("package:bnlearn",unload=TRUE)
library(igraph)
results = matrix(rep(0,6),nrow=1,byrow=TRUE)
tNames = c("dc","ec","bc","cc","prob","elrcf")
fb1=cbind(t(t(degree(gi1))),t(t(eigen_centrality(gi1)$vector)),t(t(
  betweenness(gi1))),t(t(closeness(gi1))))
fb2=cbind(t(t(degree(gi2))),t(t(eigen_centrality(gi2)$vector)),t(t(
  betweenness(gi2))),t(t(closeness(gi2))))
fb3=cbind(t(t(degree(gi3))),t(t(eigen_centrality(gi3)$vector)),t(t(
  betweenness(gi3))),t(t(closeness(gi3))))
fb4=cbind(t(t(degree(gi4))),t(t(eigen_centrality(gi4)$vector)),t(t(
  betweenness(gi4))),t(t(closeness(gi4))))

library(bnlearn)
for(i in seq(0.01,.99,.01)){
  if(100*i%%10==0)
    print(i)
  f1=populateSame(g1,i)
  f2=populateSame(g2,i)
  f3=populateSame(g3,i)
  f4=populateSame(g4,i)

  r1 = convertRiskMeasures(calculateELRAll(list(graph=g1,fit=f1)))
  r2 = convertRiskMeasures(calculateELRAll(list(graph=g2,fit=f2)))
  r3 = convertRiskMeasures(calculateELRAll(list(graph=g3,fit=f3)))
  r4 = convertRiskMeasures(calculateELRAll(list(graph=g4,fit=f4)))


  e1 = cbind(t(t(rep(i,9))),t(t(unlist(r1$elrcf))))
  e2 = cbind(t(t(rep(i,9))),t(t(unlist(r2$elrcf))))
```

```
    e3 = cbind(t(t(rep(i,9))),t(t(unlist(r3$elrcf))))
    e4 = cbind(t(t(rep(i,9))),t(t(unlist(r4$elrcf))))


    ne1 = cbind(fb1,e1)
    ne2 = cbind(fb2,e2)
    ne3 = cbind(fb3,e3)
    ne4 = cbind(fb4,e4)

    results=rbind(results,ne1,ne2,ne3,ne4)
  }
  df=data.frame(results)
  df=df[-1,]
  names(df)=tNames
  return(df)
}

testSimulation = function(){
  library(bnlearn)
  s1 = generateSupplyNetwork(100,70)
  s2 = generateSupplyNetwork(100,70)


  while(components(s1)$no!=1){
    s1=generateSupplyNetwork(100,70)
  }
  while(components(s2)$no!=1){
    s2=generateSupplyNetwork(50,34)
  }

  sh1=findAllStructuralHoles(s1)
  ta1 = sample(names(sh1),sample(5:length(names(sh1))))

  sh2=findAllStructuralHoles(s2)
  ta2 = sample(names(sh2),sample(5:length(names(sh2))))

  sh1=findAllStructuralHoles(s1)
```

```
ta1 = sample(names(sh1),sample(5:length(names(sh1))))
visited = c()
for(i in ta1){
  nums = sh1[[i]]$missing
  if(is.element(paste(nums,collapse=""),visited))
    next
  num=sapply(nums,function(x){as.numeric(substr(x,2,nchar(x)))})
  nEdge = length(E(s1))+1
  if(num[1]<num[2]){
    s1=add_edges(s1,num,attr=list(name=paste("e",nEdge,sep="")))
    nEdge = nEdge+1
    visited = c(visited,paste(nums,collapse=""),paste(rev(nums),
  collapse=""))
  }
}

r1 = generateRiskGraph(s1,10,10)
r2 = generateRiskGraph(s2,10,10)

m1 = calculateELRAll(r1$graph,r1$fit)
m2 = calculateELRAll(r2$graph,r2$fit)

m1 = convertRiskMeasures(m1)
m2 = convertRiskMeasures(m2)

r1 = c(r1,m1)
r2 = c(r2,m2)

detach("package:bnlearn",unload=TRUE)

graphData1 = list()


n1 = c(V(r1$supplyNetwork)$name,E(r1$supplyNetwork)$name)
n2 = c(V(r2$supplyNetwork)$name,E(r2$supplyNetwork)$name)

met1 =  getEgoMetrics(r1,2,n1[1])
```

```
  met2 =  getEgoMetrics(r2,2,n2[1])

  n1=n1[-1]
  n2=n2[-1]

  for(i in n1){
    met1 = rbind(met1,getEgoMetrics(r1,2,i))
  }
  for(i in n2){
    met2 = rbind(met2,getEgoMetrics(r2,2,i))
  }

  met = rbind(met1,met2)
  return(met)
}
```

Listing A.1: R Functions and code for the simulation.

Appendix B

R Code for Data Analysis

```
library(car)
library(lmtest)
library(systemfit)
library(gmm)
library(xtable)
library(psych)
library(texreg)
library(lavaan)
library(texreg)
library(ggplot2)
library(reshape2)

removeOutliers=function(data,vars=c()){
  variables = vars
  if(length(vars)==0){
      variables = names(data)
  }
  tdata = data
  for(i in variables){
    tdata = tdata[tdata[,i]>quantile(tdata[,i],.025),]
    tdata = tdata[tdata[,i]<quantile(tdata[,i],.975),]
  }
  return(tdata)
}
corstars <-function(x, method=c("pearson", "spearman"),
    removeTriangle=c("upper", "lower"),
                      result=c("none", "html", "latex")){
    #Compute correlation matrix
    require(Hmisc)
```

```
  x <- as.matrix(x)
  correlation_matrix<-rcorr(x, type=method[1])
  R <- correlation_matrix$r # Matrix of correlation coeficients
  p <- correlation_matrix$P # Matrix of p-value

  ## Define notions for significance levels; spacing is important.
  mystars <- ifelse(p < .0001, "*", ifelse(p < .001, "*", ifelse(p
 < .01, "*", ifelse(p < .05, "*", "")))

  ## trunctuate the correlation matrix to two decimal
  R <- format(round(cbind(rep(-1.11, ncol(x)), R), 2))[,-1]

  ## build a new matrix that includes the correlations with their
 apropriate stars
  Rnew <- matrix(paste(R, mystars, sep=""), ncol=ncol(x))
  diag(Rnew) <- paste(diag(R), " ", sep="")
  rownames(Rnew) <- paste(1:ncol(x),".",colnames(x),sep="")
  colnames(Rnew) <- 1:ncol(x)

  ## remove upper triangle of correlation matrix
  if(removeTriangle[1]=="upper"){
    Rnew <- as.matrix(Rnew)
    Rnew[upper.tri(Rnew, diag = TRUE)] <- ""
    Rnew <- as.data.frame(Rnew)
  }

  ## remove lower triangle of correlation matrix
  else if(removeTriangle[1]=="lower"){
    Rnew <- as.matrix(Rnew)
    Rnew[lower.tri(Rnew, diag = TRUE)] <- ""
    Rnew <- as.data.frame(Rnew)
  }

  ## remove last column and return the correlation matrix
  Rnew <- cbind(Rnew[1:length(Rnew)-1])
  if (result[1]=="none") return(Rnew)
  else{
```

```
      if(result[1]=="html") print(xtable(Rnew), type="html")
      else return(xtable(Rnew,type="latex"))
    }
}

##########################################
#Data Loading and Cleaning
##########################################

#First, load the data in:
ego2_data = read.csv("final_dissertation_data_ego2.csv")
ego_data = read.csv("final_dissertation_data.csv")
local_data = read.csv("final_dissertation_data_local.csv")

#Get rid of the indexing that are in the first two columns of each
    file:
ego2_data = ego2_data[,-c(1,2)]
ego_data = ego_data[,-c(1,2)]
local_data = local_data[,-c(1,2)]

#Get rid of duplicates in each respective data set
ego2_data = unique(ego2_data)
ego_data = unique(ego_data)
local_data = unique(local_data)

##########################################
#Exploratory Factor Analysis (EFA)
##########################################

####1-Ego Network and 2-Ego-Network Data EFA####
#Isolate the Independent Variables, The Raw Factor Data, & Dependent
    Variables from Each Data Set
iv_ego2 = ego2_data[,1:2]
rfd_ego2 = ego2_data[,3:30]
dv_ego2 = ego2_data[,31:39]
iv_risk_ego1 = ego_data[,1:9]
iv_ego1 = ego_data[,10:11]
```

```
rfd_ego1 = ego_data[,12:39]
dv_ego1 = ego_data[,40:48]

#Combine the 2-ego network and ego-network data
rfd_combined = rbind(rfd_ego2,rfd_ego1)

#Run a preliminary EFA using Minimum Residual and Principal Axis FA,
    save plot to file:
pdf("egofactorscreeplot.pdf")
factor.prelim.anal_ego = fa.parallel(rfd_combined,fm="minfes", fa="fa
    ")
dev.off()

#Next, run the actural factor analysis using Minimum Residual FA w/
    Varimax Rotation
factor.analysis_ego = fa(rfd_combined,nfactors = 5,rotate = "varimax"
    ,fm="minres")

#Write the EFA Plot to a file
pdf("egoefaplot.pdf")
fa.diagram(factor.analysis_ego,cut=0.2)
dev.off()

#Fix up the loading table to make it look nice:
factor.loadings_ego=matrix(factor.analysis_ego$loadings[1:140],ncol
    =5)
rownames(factor.loadings_ego) = dimnames(factor.analysis_ego$loadings
    )[[1]]
colnames(factor.loadings_ego) = dimnames(factor.analysis_ego$loadings
    )[[2]]
max_vals = apply(abs(factor.loadings_ego),1,max)
temp_tab = (abs(factor.loadings_ego) == max_vals)
tOrder = c()
for(i in 1:5){
  tOrder = c(tOrder,which(temp_tab[,i]))
}
factor.loadings_ego = factor.loadings_ego[tOrder,]
```

```
#Now write it to Latex Form.  This will manually be cleaned up later in the diss
print(xtable(factor.loadings_ego, type = "latex"), file = "efa_ego.tex")

#Construct the factor scores for the 2-ego-network
factor.scores_ego = factor.analysis_ego$scores
colnames(factor.scores_ego)=c("Centrality","Distance","Clustering","Feedback","Connectivity")
factor.scores_ego2 = factor.scores_ego[1:nrow(rfd_ego2),]
factor.scores_ego1 = factor.scores_ego[-(1:nrow(rfd_ego2)),]

#Combine to get the final data set for ego2 and clean it by removing outliers
iv_ego2 = cbind(iv_ego2,factor.scores_ego2)
data_ego2 = cbind(iv_ego2,dv_ego2)
data_ego2 = removeOutliers(data_ego2,c(colnames(factor.scores_ego2),names(dv_ego2)))

#Now Conduct a confirmatory factor analysis on the risk variables from ego 2:
contribution=paste("contribution=~",paste(names(iv_risk_ego1)[seq(1,9,3)],collapse="+"),sep="")
significance=paste("significance=~",paste(names(iv_risk_ego1)[seq(2,9,3)],collapse="+"),sep="")
velocity = paste("velocity=~",paste(names(iv_risk_ego1)[seq(3,9,3)],collapse="+"),sep="")
model = paste(contribution,significance,velocity,sep="\n")

iv_risk_ego1[,"cv_elrv_2ego"] = iv_risk_ego1[,"cv_elrv_2ego"]/10000
risk.factors_ego1 = cfa(model,data=iv_risk_ego1)
risk.factors_ego1.scores = predict(risk.factors_ego1)


#Combine the data for ego1
iv_ego1 = cbind(iv_ego1,risk.factors_ego1.scores,factor.scores_ego1)
```

```
data_ego1 = cbind(iv_ego1,dv_ego1)
data_ego1 = removeOutliers(data_ego1,c(colnames(factor.scores_ego1),
    names(dv_ego1)))

####Local Data EFA####

#Isolate the Independent Variables, The Raw Factor Data, & Dependent
    Variables
iv_risk_local = local_data[,3:11]
iv_local = local_data[,c(1:2,12)]
rfd_local = local_data[,13:26]
dv_local = local_data[,27:29]

#Run a preliminary EFA using Minimum Residual and Principal Axis FA,
    save plot to file:
pdf("localfactorscreeplot.pdf")
factor.prelim.anal_local = fa.parallel(rfd_local,fm="minfes", fa="fa"
    )
dev.off()

#Next, run the actural factor analysis using Minimum Residual FA w/
    Varimax Rotation
factor.analysis_local = fa(rfd_local,nfactors = 5,rotate = "varimax",
    fm="minres")

#Write the EFA Plot to a file
pdf("localefaplot.pdf")
fa.diagram(factor.analysis_local,cut=0.2)
dev.off()

#Fix up the loading table to make it look nice:
factor.loadings_local=matrix(factor.analysis_local$loadings[1:70],
    ncol=5)
rownames(factor.loadings_local) = dimnames(factor.analysis_local$
    loadings)[[1]]
colnames(factor.loadings_local) = dimnames(factor.analysis_local$
    loadings)[[2]]
```

```
max_vals = apply(abs(factor.loadings_local),1,max)
temp_tab = (abs(factor.loadings_local) == max_vals)
tOrder = c()
for(i in 1:5){
  tOrder = c(tOrder,which(temp_tab[,i]))
}
factor.loadings_local = factor.loadings_local[tOrder,]

#Now write it to Latex Form.  This will manually be cleaned up later in the diss
print(xtable(factor.loadings_local, type = "latex"), file = "efa_local.tex")

#Get the factor scores
factor.scores_local = factor.analysis_local$scores
colnames(factor.scores_local) = c("Centrality","Distance","Clustering","Feedback","Connectivity")
#Now Conduct a confirmatory factor analysis on the risk variables from ego 1:
contribution=paste("contribution=~",paste(names(iv_risk_local)[seq(1,9,3)],collapse="+"),sep="")
significance=paste("significance=~",paste(names(iv_risk_local)[seq(2,9,3)],collapse="+"),sep="")
velocity = paste("velocity=~",paste(names(iv_risk_local)[seq(3,9,3)],collapse="+"),sep="")
model = paste(contribution,significance,velocity,sep="\n")

iv_risk_local[,"cv_elrv_ego"] = iv_risk_local[,"cv_elrv_ego"]/10000
risk.factors_local = cfa(model,data=iv_risk_local)
risk.factors_local.scores = predict(risk.factors_local)

#Now combine the variables and clean them up:
iv_local = cbind(iv_local,risk.factors_local.scores,factor.scores_local)
data_local = cbind(iv_local,dv_local)
data_local = removeOutliers(data_local,c(colnames(factor.scores_local),names(dv_local)))
```

```
##########################################
#Correlation Matrices
##########################################
#For the original data-sets
ec2 = ego2_data
ec1 = ego_data
loc = local_data
#names(ec2) = 1:length(names(ec2))
#names(ec1) = 1:length(names(ec1))
#names(loc) = 1:length(names(loc))

corr_ego2_original = corstars(ec2,result="latex")
print(corr_ego2_original,file="ego2_corr_orig.tex")

corr_ego1_original = corstars(ec1,result="latex")
print(corr_ego1_original,file="ego1_corr_orig.tex")

corr_local_original = corstars(loc,result="latex")
print(corr_local_original,file="local_corr_orig.tex")

#For the factored Correlation Matrices
corr_ego2_factor = corstars(data_ego2,result="latex")
print(corr_ego2_factor,file="ego2_corr_factor.tex")

corr_ego1_factor = corstars(data_ego1,result="latex")
print(corr_ego1_factor,file="ego1_corr_factor.tex")

corr_local_factor = corstars(data_local,result="latex")
print(corr_local_factor,file="local_corr_factor.tex")


##########################################
#Plotting
##########################################
#Demonstration Plots

```

```r


#2-Ego Factor Hists
x <- data_ego2[,c(3:7)]
data<- melt(x)
ego2_hist_fac = ggplot(data,aes(x=value, fill=variable)) + geom_density(alpha=0.2)+xlab("Factor Score")+ylab("Density") + ggtitle("2-Ego-Network Factor Densities")+theme(plot.title = element_text(hjust = 0.5))+labs(fill="Factors")

#2-Ego Risk Hists
x <- data_ego2[,c(9:16)]
data<- melt(x)
ego2_hist_risk = ggplot(data,aes(x=value, fill=variable)) + geom_density(alpha=0.2)+xlab("Risk Value")+ylab("Density") + ggtitle("2-Ego-Network Risk Densities")+theme(plot.title = element_text(hjust = 0.5))+labs(fill="Risks")+xlim(-5,5)

#1-Ego Factor Hists
x <- data_ego1[,c(6:10)]
data<- melt(x)
ego1_hist_fac = ggplot(data,aes(x=value, fill=variable)) + geom_density(alpha=0.2)+xlab("Factor Score")+ylab("Density") + ggtitle("1-Ego-Network Factor Densities")+theme(plot.title = element_text(hjust = 0.5))+labs(fill="Factors")

#1-Ego Risk Hists
x <- data_ego1[,c(11:19)]
data<- melt(x)
ego1_hist_risk = ggplot(data,aes(x=value, fill=variable)) + geom_density(alpha=0.2)+xlab("Risk Value")+ylab("Density") + ggtitle("1-Ego-Network Risk Densities")+theme(plot.title = element_text(hjust = 0.5))+labs(fill="Risks")+xlim(-5,5)

#1-Ego IndRisk Hists
x <- data_ego1[,c(3:5)]
data<- melt(x)
```

```
ego1_hist_risk_ind = ggplot(data,aes(x=value, fill=variable)) + geom_density(alpha=0.2)+xlab("Independent Risk Value")+ylab("Density") + ggtitle("1-Ego-Network Independent Risk Densities")+theme(plot.title = element_text(hjust = 0.5))+labs(fill="Risks")+xlim(-5,5)


#Local Factor Hists
x <- data_local[,c(7:11)]
data<- melt(x)
local_hist_fac = ggplot(data,aes(x=value, fill=variable)) + geom_density(alpha=0.2)+xlab("Factor Score")+ylab("Density") + ggtitle("Local-Network Factor Densities")+theme(plot.title = element_text(hjust = 0.5))+labs(fill="Factors")

#Local Risk Hists
x <- data_local[,c(12:14)]
data<- melt(x)
local_hist_risk = ggplot(data,aes(x=value, fill=variable)) + geom_density(alpha=0.2)+xlab("Risk Value")+ylab("Density") + ggtitle("Local-Network Risk Densities")+theme(plot.title = element_text(hjust = 0.5))+labs(fill="Risks")+xlim(-5,5)

#Local IndRisk Hists
x <- data_local[,c(4:6)]
data<- melt(x)
local_hist_risk_ind = ggplot(data,aes(x=value, fill=variable)) + geom_density(alpha=0.2)+xlab("Independent Risk Value")+ylab("Density") + ggtitle("Local-Network Independent Risk Densities")+theme(plot.title = element_text(hjust = 0.5))+labs(fill="Risks")+xlim(-5,5)


pdf("hist1.pdf")
grid.arrange(ego2_hist_risk,ego2_hist_fac,nrow=2)
dev.off()

pdf("hist2.pdf")
```

```
grid.arrange(ego1_hist_risk_ind,ego1_hist_fac,nrow=2)
dev.off()

pdf("hist3.pdf")
grid.arrange(ego1_hist_risk,local_hist_risk,nrow=2)
dev.off()

pdf("hist4.pdf")
grid.arrange(local_hist_risk_ind,local_hist_fac,nrow=2)
dev.off()

#Final Data Set Hists by Variable
#Centrality
df = data.frame(Q = c(data_ego2$Centrality,data_ego1$Centrality,data_local$Centrality),
                type = c(rep("Ego2", nrow(data_ego2)), rep("Ego1", nrow(data_ego1)),rep("Local", nrow(data_local))))
density_cen = ggplot(df,aes(x=Q, fill=type)) + geom_density(alpha=0.7)+xlab("Factor Score")+ylab("Density") + ggtitle("Centrality Densities")+theme(plot.title = element_text(hjust = 0.5))+labs(fill="Level") + xlim(-2,2)

#Connectivity
df = data.frame(Q = c(data_ego2$Connectivity,data_ego1$Connectivity,data_local$Connectivity),
                type = c(rep("Ego2", nrow(data_ego2)), rep("Ego1", nrow(data_ego1)),rep("Local", nrow(data_local))))
density_con = ggplot(df,aes(x=Q, fill=type)) + geom_density(alpha=0.7)+xlab("Factor Score")+ylab("Density") + ggtitle("Connectivity Densities")+theme(plot.title = element_text(hjust = 0.5))+labs(fill="Level") + xlim(-2,2)

#Clustering
df = data.frame(Q = c(data_ego2$Clustering,data_ego1$Clustering,data_local$Clustering),
                type = c(rep("Ego2", nrow(data_ego2)), rep("Ego1", nrow(data_ego1)),rep("Local", nrow(data_local))))
```

```
density_clu = ggplot(df,aes(x=Q, fill=type)) + geom_density(alpha
    =0.7)+xlab("Factor Score")+ylab("Density") + ggtitle("Clustering
    Densities")+theme(plot.title = element_text(hjust = 0.5))+labs(
    fill="Level") + xlim(-2,2)

#Feedback
df = data.frame(Q = c(data_ego2$Feedback,data_ego1$Feedback,data_
    local$Feedback),
                type = c(rep("Ego2", nrow(data_ego2)), rep("Ego1",
    nrow(data_ego1)),rep("Local", nrow(data_local))))
density_fee = ggplot(df,aes(x=Q, fill=type)) + geom_density(alpha
    =0.7)+xlab("Factor Score")+ylab("Density") + ggtitle("Feedback
    Densities")+theme(plot.title = element_text(hjust = 0.5))+labs(
    fill="Level") + xlim(-2,2)

#Distance
df = data.frame(Q = c(data_ego2$Distance,data_ego1$Distance,data_
    local$Distance),
                type = c(rep("Ego2", nrow(data_ego2)), rep("Ego1",
    nrow(data_ego1)),rep("Local", nrow(data_local))))
density_dis = ggplot(df,aes(x=Q, fill=type)) + geom_density(alpha
    =0.7)+xlab("Factor Score")+ylab("Density") + ggtitle("Distance
    Densities")+theme(plot.title = element_text(hjust = 0.5))+labs(
    fill="Level") + xlim(-2,2)

pdf("level_densities.pdf")
grid.arrange(density_cen,density_con,density_clu,density_fee,density_
    dis,nrow=3)
dev.off()
#Scatter Plots
#2-Ego SPs
#Centrality
ind = cut(data_ego2$Centrality,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_ego2$cv_elrv)))
dat$cen=ind
```

```
ego2_cen_sp = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.size = -1)+ylim(-10,10)+ggtitle("2-Ego Risk v. Centrality")+ylab("Risk")+xlab("Centrality")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())
#Clustering
ind = cut(data_ego2$Clustering,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_ego2$cv_elrv)))
dat$cen=ind
ego2_cl_sp = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.size = -1)+ylim(-10,10)+ggtitle("2-Ego Risk v. Clustering")+ylab("Risk")+xlab("Clustering")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())
#Connectivity
ind = cut(data_ego2$Connectivity,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_ego2$cv_elrv)))
dat$cen=ind
ego2_con_sp = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.size = -1)+ylim(-10,10)+ggtitle("2-Ego Risk v. Connectivity")+ylab("Risk")+xlab("Connectivity")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())
#Distance
ind = cut(data_ego2$Distance,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_ego2$cv_elrv)))
dat$cen=ind
ego2_dis_sp = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.size = -1)+ylim(-10,10)+ggtitle("2-Ego Risk v. Distance")+ylab("Risk")+xlab("Distance")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())
#Feedback
ind = cut(data_ego2$Feedback,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_ego2$cv_elrv)))
dat$cen=ind
ego2_fb_sp = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.size = -1)+ylim(-10,10)+ggtitle("2-Ego Risk v. Feedback")+ylab("Risk")+xlab("Feedback")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())

```

```
pdf("ego2sp.pdf")
grid.arrange(ego2_cen_sp,ego2_cl_sp,ego2_con_sp,ego2_dis_sp,ego2_fb_
    sp,nrow = 3)
dev.off()

#1-Ego SPs
#Centrality
ind = cut(data_ego1$Centrality,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_ego1$cv_elrv)))
dat$cen=ind
ego1_cen_sp = ggplot(dat, aes(x=cen, y=risk)) +ylim(-10,10)+  geom_
    boxplot(outlier.size = -1)+ggtitle("1-Ego Risk v. Centrality")+
    ylab("Risk")+xlab("Centrality")+theme(plot.title = element_text(
    hjust = 0.5),axis.text.x=element_blank())
#Clustering
ind = cut(data_ego1$Clustering,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_ego1$cv_elrv)))
dat$cen=ind
ego1_cl_sp = ggplot(dat, aes(x=cen, y=risk)) +  ylim(-10,10)+geom_
    boxplot(outlier.size = -1)+ggtitle("1-Ego Risk v. Clustering")+
    ylab("Risk")+xlab("Clustering")+theme(plot.title = element_text(
    hjust = 0.5),axis.text.x=element_blank())
#Connectivity
ind = cut(data_ego1$Connectivity,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_ego1$cv_elrv)))
dat$cen=ind
ego1_con_sp = ggplot(dat, aes(x=cen, y=risk)) +  ylim(-10,10)+geom_
    boxplot(outlier.size = -1)+ggtitle("1-Ego Risk v. Connectivity")+
    ylab("Risk")+xlab("Connectivity")+theme(plot.title = element_text
    (hjust = 0.5),axis.text.x=element_blank())
#Distance
ind = cut(data_ego1$Distance,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_ego1$cv_elrv)))
dat$cen=ind
```

```
ego1_dis_sp = ggplot(dat, aes(x=cen, y=risk)) +  ylim(-10,10)+geom_boxplot(outlier.size = -1)+ggtitle("1-Ego Risk v. Distance")+ylab("Risk")+xlab("Distance")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())
#Feedback
ind = cut(data_ego1$Feedback,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_ego1$cv_elrv)))
dat$cen=ind
ego1_fb_sp = ggplot(dat, aes(x=cen, y=risk)) +  ylim(-10,10)+geom_boxplot(outlier.size = -1)+ggtitle("1-Ego Risk v. Feedback")+ylab("Risk")+xlab("Feedback")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())

pdf("ego1sp.pdf")
grid.arrange(ego1_cen_sp,ego1_cl_sp,ego1_con_sp,ego1_dis_sp,ego1_fb_sp,nrow = 3)
dev.off()

#1-Ego Network Risks

#Contribution
ind = cut(data_ego1$contribution,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_ego1$cv_elrv)))
dat$cen=ind
ego1_cont_sp = ggplot(dat, aes(x=cen, y=risk)) +  ylim(-10,10)+geom_boxplot(outlier.size = -1)+ggtitle("1-Ego Risk v. 2-Ego-Contri")+ylab("Risk")+xlab("Contribution")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())

#Significance
ind = cut(data_ego1$significance,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_ego1$cv_elrv)))
dat$cen=ind
ego1_sig_sp = ggplot(dat, aes(x=cen, y=risk)) +  ylim(-10,10)+geom_boxplot(outlier.size = -1)+ggtitle("1-Ego Risk v. 2-Ego-Sign")+ylab("Risk")+xlab("Significance")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())
```

```
#Velocity
ind = cut(data_ego1$velocity,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_ego1$cv_elrv)))
dat$cen=ind
ego1_vel_sp = ggplot(dat, aes(x=cen, y=risk)) +  ylim(-10,10)+geom_
    boxplot(outlier.size = -1)+ggtitle("1-Ego Risk v. 2-Ego-Vel")+
    ylab("Risk")+xlab("Velocity")+theme(plot.title = element_text(
    hjust = 0.5),axis.text.x=element_blank())

pdf("ego1spr.pdf")
grid.arrange(ego1_cont_sp,ego1_sig_sp,ego1_vel_sp,nrow = 2)
dev.off()

#Local SPs
#Centrality
ind = cut(data_local$Centrality,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_local$elrv)))
dat$cen=ind
loc_cen_sp = ggplot(dat, aes(x=cen, y=risk)) +ylim(-10,10)+  geom_
    boxplot(outlier.size = -1)+ggtitle("Local Risk v. Centrality")+
    ylab("Risk")+xlab("Centrality")+theme(plot.title = element_text(
    hjust = 0.5),axis.text.x=element_blank())
#Clustering
ind = cut(data_local$Clustering,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_local$elrv)))
dat$cen=ind
loc_cl_sp = ggplot(dat, aes(x=cen, y=risk)) +  ylim(-3,3)+geom_
    boxplot(outlier.size = -1)+ggtitle("Local Risk v. Clustering")+
    ylab("Risk")+xlab("Clustering")+theme(plot.title = element_text(
    hjust = 0.5),axis.text.x=element_blank())
#Connectivity
ind = cut(data_local$Connectivity,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_local$elrv)))
dat$cen=ind
```

```
loc_con_sp = ggplot(dat, aes(x=cen, y=risk)) +  ylim(-3,3)+geom_boxplot(outlier.size = -1)+ggtitle("Local Risk v. Connectivity")+ylab("Risk")+xlab("Connectivity")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())
#Distance
ind = cut(data_local$Distance,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_local$elrv)))
dat$cen=ind
loc_dis_sp = ggplot(dat, aes(x=cen, y=risk)) +  ylim(-3,3)+geom_boxplot(outlier.size = -1)+ggtitle("Local Risk v. Distance")+ylab("Risk")+xlab("Distance")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())
#Feedback
ind = cut(data_local$Feedback,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_local$elrv)))
dat$cen=ind
loc_fb_sp = ggplot(dat, aes(x=cen, y=risk)) +  ylim(-3,3)+geom_boxplot(outlier.size = -1)+ggtitle("Local Risk v. Feedback")+ylab("Risk")+xlab("Feedback")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())

pdf("locsp.pdf")
grid.arrange(loc_cen_sp,loc_cl_sp,loc_con_sp,loc_dis_sp,loc_fb_sp,nrow = 3)
dev.off()

#Local Network Risks

#Contribution
ind = cut(data_local$contribution,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_local$elrv)))
dat$cen=ind
loc_cont_sp = ggplot(dat, aes(x=cen, y=risk)) +  ylim(-1,2)+geom_boxplot(outlier.size = -1)+ggtitle("Local Risk v. 1-Ego-Contri")+ylab("Risk")+xlab("Contribution")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())

```

```
#Significance
ind = cut(data_local$significance,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_local$elrv)))
dat$cen=ind
loc_sig_sp = ggplot(dat, aes(x=cen, y=risk)) +  ylim(-1,2)+geom_
    boxplot(outlier.size = -1)+ggtitle("Local Risk v. 1-Ego-Sign")+
    ylab("Risk")+xlab("Significance")+theme(plot.title = element_text
    (hjust = 0.5),axis.text.x=element_blank())

#Velocity
ind = cut(data_local$velocity,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(data_local$elrv)))
dat$cen=ind
loc_vel_sp = ggplot(dat, aes(x=cen, y=risk)) +  ylim(-1,2)+geom_
    boxplot(outlier.size = -1)+ggtitle("Local Risk v. 1-Ego-Vel")+
    ylab("Risk")+xlab("Velocity")+theme(plot.title = element_text(
    hjust = 0.5),axis.text.x=element_blank())

pdf("locspr.pdf")
grid.arrange(loc_cont_sp,loc_sig_sp,loc_vel_sp,nrow = 2)
dev.off()

#Final Functions
#2-Ego-Network Risks
#pdf("ego2functions.pdf")
p1=ggplot(data.frame(x = c(0, 1)), aes(x = x)) +
  stat_function(fun=function(x).39*x^2-.01*x+2.01, aes(colour="
    Connectivity"),size=1.5)+
stat_function(fun=function(x) -.1*x^2-.07*x+2.01, aes(colour="
    Feedback"),size=1.5)+
  stat_function(fun=function(x)-.08*x^2+.35*x+2.01, aes(colour="
    Clustering"),size=1.5)+
  stat_function(fun=function(x).08*x^2-.71*x+2.01, aes(colour="
    Distance"),size=1.5)+
stat_function(fun=function(x)-.24*x^2+1.09*x+2.01, aes(colour="
    Centrality"),size=1.5)+
  scale_x_continuous(name = "Factor Value",
```

```
                      breaks = seq(-2,2,.5),
                      limits=c(-2, 2)) +
  scale_y_continuous(name = "2-Ego-Network Risk") +
  ggtitle("2-Ego-Network Risk") +
  scale_colour_manual("Variables", values = c("#4271AE", "#1F3552","
    darkslateblue","firebrick","green4")) +
  theme_economist()+
  theme(legend.position = "bottom", legend.direction = "horizontal",
        legend.box = "horizontal",
        legend.key.size = unit(1, "cm"),
        plot.title = element_text(family="Helvetica"),
        text = element_text(family = "Helvetica"),
        axis.title = element_text(size = 12),
        legend.text = element_text(size = 9),
        legend.title=element_text(face = "bold", size = 9))

#dev.off()

#1-Ego-Network Risks
#pdf("ego1functions.pdf")
p2=ggplot(data.frame(x = c(0, 1)), aes(x = x)) +
  stat_function(fun=function(x) -.04*x+1.68, aes(colour="Connectivity
    "),size=1.5)+
  stat_function(fun=function(x) -.2*x^2-.05*x+1.68, aes(colour="
    Feedback"),size=1.5)+
  stat_function(fun=function(x) -.18*x^2+.63*x+1.68, aes(colour="
    Clustering"),size=1.5)+
  stat_function(fun=function(x) .11*x^2+.63*x+1.68, aes(colour="
    Distance"),size=1.5)+
  stat_function(fun=function(x) .89*x^2+1.13*x+1.68, aes(colour="
    Centrality"),size=1.5)+
  scale_x_continuous(name = "Factor Value",
                      breaks = seq(-2,2,.5),
                      limits=c(-2, 2)) +
  scale_y_continuous(name = "1-Ego-Network Risk") +
  ggtitle("1-Ego-Network Risk") +
```

```
  scale_colour_manual("Variables", values = c("#4271AE", "#1F3552","
    darkslateblue","firebrick","green4")) +
  theme_economist()+
  theme(legend.position = "bottom", legend.direction = "horizontal",
        legend.box = "horizontal",
        legend.key.size = unit(1, "cm"),
        plot.title = element_text(family="Helvetica"),
        text = element_text(family = "Helvetica"),
        axis.title = element_text(size = 12),
        legend.text = element_text(size = 9),
        legend.title=element_text(face = "bold", size = 9))

#dev.off()
#1-Ego-Network Risks, Ind-Risks
#pdf("ego1rfunctions.pdf")
p3=ggplot(data.frame(x = c(0, 1)), aes(x = x)) +
  stat_function(fun=function(x) .1*x+1.68, aes(colour="2-Ego Velocity
    "),size=1.5)+
  stat_function(fun=function(x) -.15*x+1.68, aes(colour="2-Ego
    Significance"),size=1.5)+
  stat_function(fun=function(x) .01*x+1.68, aes(colour="2-Ego
    Contribution"),size=1.5)+
  scale_x_continuous(name = "Factor Value",
                     breaks = seq(-10,10,2),
                     limits=c(-10, 10)) +
  scale_y_continuous(name = "1-Ego-Network Risk") +
  ggtitle("1-Ego-Network Risk vs 2-Ego-Network-Risk") +
  scale_colour_manual("Variables", values = c("#4271AE", "#1F3552","
    darkslateblue","firebrick","green4")) +
  theme_economist()+
  theme(legend.position = "bottom", legend.direction = "horizontal",
        legend.box = "horizontal",
        legend.key.size = unit(1, "cm"),
        plot.title = element_text(family="Helvetica"),
        text = element_text(family = "Helvetica"),
        axis.title = element_text(size = 12),
        legend.text = element_text(size = 9),
```

```
        legend.title=element_text(face = "bold", size = 9))

#dev.off()
#Local- Risks
#pdf("localfunctions.pdf")
p4=ggplot(data.frame(x = c(0, 1)), aes(x = x)) +
  stat_function(fun=function(x) .16*x^2-.42*x-.43, aes(colour="Connectivity"),size=1.5)+
  stat_function(fun=function(x) -.03*x^2-.09*x-.43, aes(colour="Feedback"),size=1.5)+
  stat_function(fun=function(x) .14*x-.43, aes(colour="Clustering"), size=1.5)+
  stat_function(fun=function(x) .01*x^2-.18*x-.43, aes(colour="Distance"),size=1.5)+
  stat_function(fun=function(x) .02*x^2+.29*x-.43, aes(colour="Centrality"),size=1.5)+
  scale_x_continuous(name = "Factor Value",
                     breaks = seq(-2,2,.5),
                     limits=c(-2, 2)) +
  scale_y_continuous(name = "Local Risk") +
  ggtitle("Local Risk") +
  scale_colour_manual("Variables", values = c("#4271AE", "#1F3552","darkslateblue","firebrick","green4")) +
  theme_economist()+
  theme(legend.position = "bottom", legend.direction = "horizontal",
        legend.box = "horizontal",
        legend.key.size = unit(1, "cm"),
        plot.title = element_text(family="Helvetica"),
        text = element_text(family = "Helvetica"),
        axis.title = element_text(size = 12),
        legend.text = element_text(size = 9),
        legend.title=element_text(face = "bold", size = 9))

#dev.off()

#Local- Risks
#pdf("localrfunctions.pdf")
```

```
p5=ggplot(data.frame(x = c(0, 1)), aes(x = x)) +
  stat_function(fun=function(x) .01*x-.43, aes(colour="1-Ego Velocity
    "),size=1.5)+
  stat_function(fun=function(x) -.02*x-.43, aes(colour="1-Ego
    Significance"),size=1.5)+
  stat_function(fun=function(x) -.43, aes(colour="1-Ego Contribution"
    ),size=1.5)+
  scale_x_continuous(name = "Factor Value",
                     breaks = seq(-10,10,2),
                     limits=c(-10, 10)) +
  scale_y_continuous(name = "Local Risk") +
  ggtitle("Local Risk vs 1-Ego-Risk") +
  scale_colour_manual("Variables", values = c("#4271AE", "#1F3552","
    darkslateblue","firebrick","green4")) +
  theme_economist()+
  theme(legend.position = "bottom", legend.direction = "horizontal",
        legend.box = "horizontal",
        legend.key.size = unit(1, "cm"),
        plot.title = element_text(family="Helvetica"),
        text = element_text(family = "Helvetica"),
        axis.title = element_text(size = 12),
        legend.text = element_text(size = 9),
        legend.title=element_text(face = "bold", size = 9))

#dev.off()
pdf("grid.pdf")
grid.arrange(p1,p2,p3,p4,nrow = 2)
dev.off()

#Demonstration 1 Plots
df = demonstration()

#diameter
ind = cut(df$diameter,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(df$aelrcf)))
dat$cen=ind
```

```
dem1_diam = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.size = -1)+ylim(0,5)+ggtitle("Risk v. Diameter")+ylab("Avg. ELRCF")+xlab("Diameter")+theme(plot.title = element_text(hjust = 0.5), axis.text.x=element_blank())

#radius
ind = cut(df$radius,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(df$aelrcf)))
dat$cen=ind
dem1_rad = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.size = -1)+ylim(0,5)+ggtitle("Risk v. Radius")+ylab("Avg. ELRCF")+xlab("Radius")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())

#acc
ind = cut(df$acc,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(df$aelrcf)))
dat$cen=ind
dem1_acc = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.size = -1)+ylim(0,5)+ggtitle("Risk v. Closeness Cen")+ylab("Avg. ELRCF")+xlab("Avg Closeness Centrality")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())

#abc
ind = cut(df$abc,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(df$aelrcf)))
dat$cen=ind
dem1_abc = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.size = -1)+ylim(0,5)+ggtitle("Risk v. Betweenness Cen")+ylab("Avg. ELRCF")+xlab("Avg Betweenness Centrality")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())

#aevc
ind = cut(df$aec,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(df$aelrcf)))
dat$cen=ind
```

```
dem1_aec = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.size = -1)+ylim(0,5)+ggtitle("Risk v. Eigen Cen")+ylab("Avg. ELRCF")+xlab("Avg Eigenevector Centrality")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())

pdf("avgrisk.pdf")
grid.arrange(dem1_abc,dem1_acc,dem1_aec,dem1_diam,dem1_rad,nrow = 2)
dev.off()

#diameter
ind = cut(df$diameter,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(df$cvelrcf)))
dat$cen=ind
dem1_diam = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.size = -1)+ylim(.5,1)+ggtitle("Risk v. Diameter")+ylab("Avg. ELRCF")+xlab("Diameter")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())

#radius
ind = cut(df$radius,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(df$cvelrcf)))
dat$cen=ind
dem1_rad = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.size = -1)+ylim(.5,1)+ggtitle("Risk v. Radius")+ylab("Avg. ELRCF")+xlab("Radius")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())

#acc
ind = cut(df$acc,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(df$cvelrcf)))
dat$cen=ind
dem1_acc = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.size = -1)+ylim(.5,1)+ggtitle("Risk v. Closeness Cen")+ylab("Avg. ELRCF")+xlab("Avg Closeness Centrality")+theme(plot.title = element_text(hjust = 0.5),axis.text.x=element_blank())

#abc
```

```
ind = cut(df$abc,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(df$cvelrcf)))
dat$cen=ind
dem1_abc = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.
    size = -1)+ylim(.5,1)+ggtitle("Risk v. Betweenness Cen")+ylab("
    Avg. ELRCF")+xlab("Avg Betweenness Centrality")+theme(plot.title
    = element_text(hjust = 0.5),axis.text.x=element_blank())

#aevc
ind = cut(df$aec,breaks=10)
dat = data.frame(cbind(ind,risk = as.numeric(df$cvelrcf)))
dat$cen=ind
dem1_aec = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.
    size = -1)+ylim(.5,1)+ggtitle("Risk v. Eigen Cen")+ylab("Avg.
    ELRCF")+xlab("Avg Eigenevector Centrality")+theme(plot.title =
    element_text(hjust = 0.5),axis.text.x=element_blank())

pdf("cvrisk.pdf")
grid.arrange(dem1_abc,dem1_acc,dem1_aec,dem1_diam,dem1_rad,nrow = 2)
dev.off()

df = demonstration2()
#diameter
ind = cut(df$dc,breaks=20)
dat = data.frame(cbind(ind,risk = as.numeric(df$elrcf)))
dat$cen=ind
dem1_diam = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.
    size = -1)+ggtitle("Risk v. Degree Cen")+ylab("Avg. ELRCF")+xlab(
    "Degree Cen")+theme(plot.title = element_text(hjust = 0.5),axis.
    text.x=element_blank())

#radius
ind = cut(df$ec,breaks=20)
dat = data.frame(cbind(ind,risk = as.numeric(df$elrcf)))
dat$cen=ind
```

```
dem1_rad = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.
    size = -1)+ggtitle("Risk v. Eigen Cen")+ylab("Avg. ELRCF")+xlab("
    Eigen Cen")+theme(plot.title = element_text(hjust = 0.5),axis.
    text.x=element_blank())

#acc
ind = cut(df$bc,breaks=20)
dat = data.frame(cbind(ind,risk = as.numeric(df$elrcf)))
dat$cen=ind
dem1_acc = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.
    size = -1)+ggtitle("Risk v. Betweenness Cen")+ylab("Avg. ELRCF")+
    xlab("Betweenness Cen")+theme(plot.title = element_text(hjust =
    0.5),axis.text.x=element_blank())

#abc
ind = cut(df$cc,breaks=20)
dat = data.frame(cbind(ind,risk = as.numeric(df$elrcf)))
dat$cen=ind
dem1_abc = ggplot(dat, aes(x=cen, y=risk)) +  geom_boxplot(outlier.
    size = -1)+ggtitle("Risk v. Closeness Cen")+ylab("Avg. ELRCF")+
    xlab("Closeness Centrality")+theme(plot.title = element_text(
    hjust = 0.5),axis.text.x=element_blank())

pdf("cvrisklocal.pdf")
grid.arrange(dem1_abc,dem1_acc,dem1_diam,dem1_rad,nrow = 2)
dev.off()
##########################################
#Summary Statistics
##########################################

#Summary Statistics for original data set
stargazer(ego2_data)
stargazer(ego_data)
stargazer(local_data)

stargazer(data_ego2)
stargazer(data_ego1)
```

```
stargazer(data_local)

stargazer(networkChar[,-1])
##########################################################
#################The Data Analysis#########################
##########################################################

##########################################
#Some data prep
##########################################
#Get the squares of the data
local_sq_names = names(data_local)[7:11]
ego1_sq_names = names(data_ego1)[6:10]
ego2_sq_names = names(data_ego2)[3:7]

local_sq_names2=paste(local_sq_names,"2",sep="")
ego1_sq_names2=paste(ego1_sq_names,"2",sep="")
ego2_sq_names2=paste(ego2_sq_names,"2",sep="")

local_sq = data_local[,local_sq_names]^2
ego1_sq = data_ego1[,ego1_sq_names]^2
ego2_sq = data_ego2[,ego2_sq_names]^2

colnames(local_sq) = local_sq_names2
colnames(ego1_sq) = ego1_sq_names2
colnames(ego2_sq) = ego2_sq_names2

#Primary datasets
#Node-Only Data
data_node = data_local[data_local[,3]==1,-3]
#Edge-Only Data
data_edge = data_local[data_local[,3]==0,-3]

#Pooled Node/Edge Data
data_local = cbind(data_local,local_sq)
#1-Ego-Network Data
data_ego1 = cbind(data_ego1,ego1_sq)
```

```
#2-Ego-Network Data
data_ego2 = cbind(data_ego2,ego2_sq)

#All-Pooled Data
trows = nrow(data_ego1)+nrow(data_ego2)+nrow(data_local)
tcols = ncol(data_ego1)+ncol(data_ego2)+ncol(data_local)

m = matrix(rep(0,trows*tcols),nrow = trows)

rrow = 1
rcol = 22
for(i in 1:nrow(data_ego2)){
  rcol=1
  for(j in 1:ncol(data_ego2)){
    m[rrow,rcol] = data_ego2[i,j]
    rcol=rcol+1
  }
  rrow=rrow+1
}
print("Ego1")
for(i in 1:nrow(data_ego1)){
  rcol=22
  for(j in 1:ncol(data_ego1)){
    m[rrow,rcol] = data_ego1[i,j]
    rcol=rcol+1
  }
  rrow=rrow+1
}
print("Local")
for(i in 1:nrow(data_local)){
  rcol=46
  for(j in 1:ncol(data_local)){
    m[rrow,rcol] = data_local[i,j]
    rcol=rcol+1
  }
  rrow=rrow+1
}
```

```
colnames(m) = c(paste("ego2_",colnames(data_ego2),sep=""),paste("ego1_",colnames(data_ego1),sep=""),paste("local_",colnames(data_local),sep=""))
data_pooled = m
alpha = c()
beta = c()

for(i in 1:nrow(data_pooled)){
  alpha = max(data_pooled[i,c("ego2_alpha","ego1_alpha","local_alpha")])
  beta = max(data_pooled[i,c("ego2_beta","ego1_beta","local_beta")])
}

data_pooled = data.frame(cbind(alpha,beta,data_pooled[,-c(1,2,22,23,46,47)]))
#Construct the various model text to make life easy :), these are for non-pooled data
#Linear Models for Ego2
lace2 = avg_elrcf~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity
lase2 = avg_elrsf~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity
lave2 = avg_elrv~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity

lsce2 = sd_elrcf~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity
lsse2 = sd_elrsf~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity
lsve2 = sd_elrv~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity

lcce2 = cv_elrcf~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity
lcse2 = cv_elrsf~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity
```

```
lcve2 = cv_elrv~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity


#Linear Models for Ego1
lace1 = avg_elrcf~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity
lase1 = avg_elrsf~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity
lave1 = avg_elrv~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity

lsce1 = sd_elrcf~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity
lsse1 = sd_elrsf~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity
lsve1 = sd_elrv~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity

lcce1 = cv_elrcf~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity
lcse1 = cv_elrsf~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity
lcve1 = cv_elrv~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity

#Linear Models for Ego1 (Non-Pooled Node and Edge)
lcl = elrcf~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity
lsl = elrsf~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity
lvl = elrv~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity

#Linear Models for Ego1 (Pooled Node and Edge)
lclp = elrcf~alpha+beta+isNode+Centrality+Distance+Clustering+
    Feedback+Connectivity+contribution+significance+velocity
```

```
lslp = elrsf~alpha+beta+isNode+Centrality+Distance+Clustering+Feedback+Connectivity+contribution+significance+velocity
lvlp = elrv~alpha+beta+isNode+Centrality+Distance+Clustering+Feedback+Connectivity+contribution+significance+velocity

#Curvi-Linear Models for Ego2
clace2 = avg_elrcf~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity+Centrality2+Distance2+Clustering2+Feedback2+Connectivity2
clase2 = avg_elrsf~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity+Centrality2+Distance2+Clustering2+Feedback2+Connectivity2
clave2 = avg_elrv~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity+Centrality2+Distance2+Clustering2+Feedback2+Connectivity2

clsce2 = sd_elrcf~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity+Centrality2+Distance2+Clustering2+Feedback2+Connectivity2
clsse2 = sd_elrsf~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity+Centrality2+Distance2+Clustering2+Feedback2+Connectivity2
clsve2 = sd_elrv~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity+Centrality2+Distance2+Clustering2+Feedback2+Connectivity2

clcce2 = cv_elrcf~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity+Centrality2+Distance2+Clustering2+Feedback2+Connectivity2
clcse2 = cv_elrsf~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity+Centrality2+Distance2+Clustering2+Feedback2+Connectivity2
clcve2 = cv_elrv~alpha+beta+Centrality+Distance+Clustering+Feedback+Connectivity+Centrality2+Distance2+Clustering2+Feedback2+Connectivity2

#Curvi-Linear Models for Ego1
```

```
clace1 = avg_elrcf~alpha+beta+Centrality+Distance+Clustering+Feedback
    +Connectivity+contribution+significance+velocity+Centrality2+
    Distance2+Clustering2+Feedback2+Connectivity2
clase1 = avg_elrsf~alpha+beta+Centrality+Distance+Clustering+Feedback
    +Connectivity+contribution+significance+velocity+Centrality2+
    Distance2+Clustering2+Feedback2+Connectivity2
clave1 = avg_elrv~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity+Centrality2+
    Distance2+Clustering2+Feedback2+Connectivity2

clsce1 = sd_elrcf~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity+Centrality2+
    Distance2+Clustering2+Feedback2+Connectivity2
clsse1 = sd_elrsf~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity+Centrality2+
    Distance2+Clustering2+Feedback2+Connectivity2
clsve1 = sd_elrv~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity+Centrality2+
    Distance2+Clustering2+Feedback2+Connectivity2

clcce1 = cv_elrcf~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity+Centrality2+
    Distance2+Clustering2+Feedback2+Connectivity2
clcse1 = cv_elrsf~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity+Centrality2+
    Distance2+Clustering2+Feedback2+Connectivity2
clcve1 = cv_elrv~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity+Centrality2+
    Distance2+Clustering2+Feedback2+Connectivity2

#Curvi-Linear Models for Ego1 (Non-Pooled Node and Edge)
clcl = elrcf~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity+Centrality2+
    Distance2+Clustering2+Feedback2+Connectivity2
clsl = elrsf~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity+Centrality2+
    Distance2+Clustering2+Feedback2+Connectivity2
```

```
clvl = elrv~alpha+beta+Centrality+Distance+Clustering+Feedback+
    Connectivity+contribution+significance+velocity+Centrality2+
    Distance2+Clustering2+Feedback2+Connectivity2

#Curvi-Linear Models for Ego1 (Pooled Node and Edge)
clclp = elrcf~alpha+beta+isNode+Centrality+Distance+Clustering+
    Feedback+Connectivity+contribution+significance+velocity+
    Centrality2+Distance2+Clustering2+Feedback2+Connectivity2
clslp = elrsf~alpha+beta+isNode+Centrality+Distance+Clustering+
    Feedback+Connectivity+contribution+significance+velocity+
    Centrality2+Distance2+Clustering2+Feedback2+Connectivity2
clvlp = elrv~alpha+beta+isNode+Centrality+Distance+Clustering+
    Feedback+Connectivity+contribution+significance+velocity+
    Centrality2+Distance2+Clustering2+Feedback2+Connectivity2

#Pooled Data Sets:
#Construct the various model text to make life easy :)

#Linear Models for Ego2
place2 = ego2_avg_elrcf~alpha+beta+ego2_Centrality+ego2_Distance+ego2
    _Clustering+ego2_Feedback+ego2_Connectivity
plase2 = ego2_avg_elrsf~alpha+beta+ego2_Centrality+ego2_Distance+ego2
    _Clustering+ego2_Feedback+ego2_Connectivity
plave2 = ego2_avg_elrv~alpha+beta+ego2_Centrality+ego2_Distance+ego2_
    Clustering+ego2_Feedback+ego2_Connectivity

plsce2 = ego2_sd_elrcf~alpha+beta+ego2_Centrality+ego2_Distance+ego2_
    Clustering+ego2_Feedback+ego2_Connectivity
plsse2 = ego2_sd_elrsf~alpha+beta+ego2_Centrality+ego2_Distance+ego2_
    Clustering+ego2_Feedback+ego2_Connectivity
plsve2 = ego2_sd_elrv~alpha+beta+ego2_Centrality+ego2_Distance+ego2_
    Clustering+ego2_Feedback+ego2_Connectivity

plcce2 = ego2_cv_elrcf~alpha+beta+ego2_Centrality+ego2_Distance+ego2_
    Clustering+ego2_Feedback+ego2_Connectivity
plcse2 = ego2_cv_elrsf~alpha+beta+ego2_Centrality+ego2_Distance+ego2_
    Clustering+ego2_Feedback+ego2_Connectivity
```

```
plcve2 = ego2_cv_elrv~alpha+beta+ego2_Centrality+ego2_Distance+ego2_Clustering+ego2_Feedback+ego2_Connectivity

#Linear Models for Ego1
place1 = ego1_avg_elrcf~alpha+beta+ego1_Centrality+ego1_Distance+ego1_Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+ego1_significance+ego1_velocity
plase1 = ego1_avg_elrsf~alpha+beta+ego1_Centrality+ego1_Distance+ego1_Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+ego1_significance+ego1_velocity
plave1 = ego1_avg_elrv~alpha+beta+ego1_Centrality+ego1_Distance+ego1_Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+ego1_significance+ego1_velocity

plsce1 = ego1_sd_elrcf~alpha+beta+ego1_Centrality+ego1_Distance+ego1_Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+ego1_significance+ego1_velocity
plsse1 = ego1_sd_elrsf~alpha+beta+ego1_Centrality+ego1_Distance+ego1_Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+ego1_significance+ego1_velocity
plsve1 = ego1_sd_elrv~alpha+beta+ego1_Centrality+ego1_Distance+ego1_Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+ego1_significance+ego1_velocity

plcce1 = ego1_cv_elrcf~alpha+beta+ego1_Centrality+ego1_Distance+ego1_Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+ego1_significance+ego1_velocity
plcse1 = ego1_cv_elrsf~alpha+beta+ego1_Centrality+ego1_Distance+ego1_Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+ego1_significance+ego1_velocity
plcve1 = ego1_cv_elrv~alpha+beta+ego1_Centrality+ego1_Distance+ego1_Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+ego1_significance+ego1_velocity

#Linear Models for Ego1 (Non-Pooled Node and Edge)
```

```
plcl = local_elrcf~alpha+beta+local_Centrality+local_Distance+local_Clustering+local_Feedback+local_Connectivity+local_contribution+local_significance+local_velocity
plsl = local_elrsf~alpha+beta+local_Centrality+local_Distance+local_Clustering+local_Feedback+local_Connectivity+local_contribution+local_significance+local_velocity
plvl = local_elrv~alpha+beta+local_Centrality+local_Distance+local_Clustering+local_Feedback+local_Connectivity+local_contribution+local_significance+local_velocity

#Linear Models for Ego1 (Pooled Node and Edge)
plclp = local_elrcf~alpha+beta+local_isNode+local_Centrality+local_Distance+local_Clustering+local_Feedback+local_Connectivity+local_contribution+local_significance+local_velocity
plslp = local_elrsf~alpha+beta+local_isNode+local_Centrality+local_Distance+local_Clustering+local_Feedback+local_Connectivity+local_contribution+local_significance+local_velocity
plvlp = local_elrv~alpha+beta+local_isNode+local_Centrality+local_Distance+local_Clustering+local_Feedback+local_Connectivity+local_contribution+local_significance+local_velocity

#Curvi-Linear Models for Ego2
pclace2 = ego2_avg_elrcf~alpha+beta+ego2_Centrality+ego2_Distance+ego2_Clustering+ego2_Feedback+ego2_Connectivity+ego2_Centrality2+ego2_Distance2+ego2_Clustering2+ego2_Feedback2+ego2_Connectivity2
pclase2 = ego2_avg_elrsf~alpha+beta+ego2_Centrality+ego2_Distance+ego2_Clustering+ego2_Feedback+ego2_Connectivity+ego2_Centrality2+ego2_Distance2+ego2_Clustering2+ego2_Feedback2+ego2_Connectivity2
pclave2 = ego2_avg_elrv~alpha+beta+ego2_Centrality+ego2_Distance+ego2_Clustering+ego2_Feedback+ego2_Connectivity+ego2_Centrality2+ego2_Distance2+ego2_Clustering2+ego2_Feedback2+ego2_Connectivity2

pclsce2 = ego2_sd_elrcf~alpha+beta+ego2_Centrality+ego2_Distance+ego2_Clustering+ego2_Feedback+ego2_Connectivity+ego2_Centrality2+ego2_Distance2+ego2_Clustering2+ego2_Feedback2+ego2_Connectivity2
```

```
pclsse2 = ego2_sd_elrsf~alpha+beta+ego2_Centrality+ego2_Distance+ego2
    _Clustering+ego2_Feedback+ego2_Connectivity+ego2_Centrality2+ego2
    _Distance2+ego2_Clustering2+ego2_Feedback2+ego2_Connectivity2
pclsve2 = ego2_sd_elrv~alpha+beta+ego2_Centrality+ego2_Distance+ego2_
    Clustering+ego2_Feedback+ego2_Connectivity+ego2_Centrality2+ego2_
    Distance2+ego2_Clustering2+ego2_Feedback2+ego2_Connectivity2

pclcce2 = ego2_cv_elrcf~alpha+beta+ego2_Centrality+ego2_Distance+ego2
    _Clustering+ego2_Feedback+ego2_Connectivity+ego2_Centrality2+ego2
    _Distance2+ego2_Clustering2+ego2_Feedback2+ego2_Connectivity2
pclcse2 = ego2_cv_elrsf~alpha+beta+ego2_Centrality+ego2_Distance+ego2
    _Clustering+ego2_Feedback+ego2_Connectivity+ego2_Centrality2+ego2
    _Distance2+ego2_Clustering2+ego2_Feedback2+ego2_Connectivity2
pclcve2 = ego2_cv_elrv~alpha+beta+ego2_Centrality+ego2_Distance+ego2_
    Clustering+ego2_Feedback+ego2_Connectivity+ego2_Centrality2+ego2_
    Distance2+ego2_Clustering2+ego2_Feedback2+ego2_Connectivity2

#Curvi-Linear Models for Ego1
pclace1 = ego1_avg_elrcf~alpha+beta+ego1_Centrality+ego1_Distance+
    ego1_Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution
    +ego1_significance+ego1_velocity+ego1_Centrality2+ego1_Distance2+
    ego1_Clustering2+ego1_Feedback2+ego1_Connectivity2
pclase1 = ego1_avg_elrsf~alpha+beta+ego1_Centrality+ego1_Distance+
    ego1_Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution
    +ego1_significance+ego1_velocity+ego1_Centrality2+ego1_Distance2+
    ego1_Clustering2+ego1_Feedback2+ego1_Connectivity2
pclave1 = ego1_avg_elrv~alpha+beta+ego1_Centrality+ego1_Distance+ego1
    _Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+
    ego1_significance+ego1_velocity+ego1_Centrality2+ego1_Distance2+
    ego1_Clustering2+ego1_Feedback2+ego1_Connectivity2

pclsce1 = ego1_sd_elrcf~alpha+beta+ego1_Centrality+ego1_Distance+ego1
    _Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+
    ego1_significance+ego1_velocity+ego1_Centrality2+ego1_Distance2+
    ego1_Clustering2+ego1_Feedback2+ego1_Connectivity2
```

```
pclsse1 = ego1_sd_elrsf~alpha+beta+ego1_Centrality+ego1_Distance+ego1
    _Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+
    ego1_significance+ego1_velocity+ego1_Centrality2+ego1_Distance2+
    ego1_Clustering2+ego1_Feedback2+ego1_Connectivity2
pclsve1 = ego1_sd_elrv~alpha+beta+ego1_Centrality+ego1_Distance+ego1_
    Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+ego1
    _significance+ego1_velocity+ego1_Centrality2+ego1_Distance2+ego1_
    Clustering2+ego1_Feedback2+ego1_Connectivity2

pclcce1 = ego1_cv_elrcf~alpha+beta+ego1_Centrality+ego1_Distance+ego1
    _Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+
    ego1_significance+ego1_velocity+ego1_Centrality2+ego1_Distance2+
    ego1_Clustering2+ego1_Feedback2+ego1_Connectivity2
pclcse1 = ego1_cv_elrsf~alpha+beta+ego1_Centrality+ego1_Distance+ego1
    _Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+
    ego1_significance+ego1_velocity+ego1_Centrality2+ego1_Distance2+
    ego1_Clustering2+ego1_Feedback2+ego1_Connectivity2
pclcve1 = ego1_cv_elrv~alpha+beta+ego1_Centrality+ego1_Distance+ego1_
    Clustering+ego1_Feedback+ego1_Connectivity+ego1_contribution+ego1
    _significance+ego1_velocity+ego1_Centrality2+ego1_Distance2+ego1_
    Clustering2+ego1_Feedback2+ego1_Connectivity2

#Curvi-Linear Models for Ego1 (Pooled Node and Edge)
pclclp = local_elrcf~alpha+ beta+ local_isNode+ local_Centrality+
    local_Distance+ local_Clustering+ local_Feedback+ local_
    Connectivity+ local_contribution+ local_significance+ local_
    velocity+ local_Centrality2+ local_Distance2+ local_Clustering2+
    local_Feedback2+ local_Connectivity2
pclslp = local_elrsf~alpha+ beta+ local_isNode+ local_Centrality+
    local_Distance+ local_Clustering+ local_Feedback+ local_
    Connectivity+ local_contribution+ local_significance+ local_
    velocity+ local_Centrality2+ local_Distance2+ local_Clustering2+
    local_Feedback2+ local_Connectivity2
```

```
pclvlp = local_elrv~alpha+ beta+ local_isNode+ local_Centrality+
    local_Distance+ local_Clustering+ local_Feedback+ local_
    Connectivity+ local_contribution+ local_significance+ local_
    velocity+ local_Centrality2+ local_Distance2+ local_Clustering2+
    local_Feedback2+ local_Connectivity2

##########################################
#System Linear Regression for Each Level
#(Linear Model)
##########################################

####2-Ego-Network Structure
#SUR
lmsurego2 = systemfit(list(lace2,lsce2,lcce2,lase2,lsse2,lcse2,lave2,
    lsve2,lcve2),method="SUR",data=data_ego2)

####1-Ego-Network Structure
#SUR
lmsurego1 = systemfit(list(lace1,lsce1,lcce1,lase1,lsse1,lcse1,lave1,
    lsve1,lcve1),method="SUR",data=data_ego1)

####Local-Network Structure
#SUR
lmsurelocal = systemfit(list(lclp,lslp,lvlp),method="SUR",data=data_
    local)


##########################################
#System Linear Regression for Each Level
#(Curvi-Linear Model)
##########################################

####2-Ego-Network Structure
#SUR
clmsurego2 = systemfit(list(clace2,clsce2,clcce2,clase2,clsse2,clcse2
    ,clave2,clsve2,clcve2),method="SUR",data=data_ego2)
#Format these nice:
```

```
rnames = c()
sigmas = c()
r2 = c()
ar2 = c()
x=c()
for(i in clmsurego2$eq){
  si = summary(i)
  coef = si$coefficients
  rnames = rownames(coef)
  sig = coef[,4]
  sig = sapply(sig,function(x){ifelse(x<.05,"*","")})
  co = round(coef[,1],2)
  toBind2 = paste("(",round(coef[,2],2),")",sep="")
  toBind = paste(co,toBind2,sig,sep="")
  x=cbind(x,toBind)
  r2=c(r2,si$r.squared)
  ar2 = c(ar2,si$adj.r.squared)
  sigmas = c(sigmas,si$sigma)
}
mnames = c("mu_c","mu_s","mu_v","sigma_c","sigma_s","sigma_v","c_c","
    c_s","c_v")
colnames(x) = mnames
x = rbind(x,round(sigmas,2))
x = rbind(x,round(r2,2))
x = rbind(x,round(ar2,2))
rownames(x) = c(rnames,"sigma","R2","Adjusted R2")
ego2regout = x
####1-Ego-Network Structure
#SUR
clmsurego1 = systemfit(list(clace1,clsce1,clcce1,clase1,clsse1,clcse1
    ,clave1,clsve1,clcve1),method="SUR",data=data_ego1)
#Format these nice:
rnames = c()
sigmas = c()
r2 = c()
ar2 = c()
x=c()
```

```
for(i in clmsurego1$eq){
  si = summary(i)
  coef = si$coefficients
  rnames = rownames(coef)
  sig = coef[,4]
  sig = sapply(sig,function(x){ifelse(x<.05,"*","")})
  co = round(coef[,1],2)
  toBind2 = paste("(",round(coef[,2],2),")",sep="")
  toBind = paste(co,toBind2,sig,sep="")
  x=cbind(x,toBind)
  r2=c(r2,si$r.squared)
  ar2 = c(ar2,si$adj.r.squared)
  sigmas = c(sigmas,si$sigma)
}
mnames = c("mu_c","mu_s","mu_v","sigma_c","sigma_s","sigma_v","c_c","
    c_s","c_v")
colnames(x) = mnames
x = rbind(x,round(sigmas,2))
x = rbind(x,round(r2,2))
x = rbind(x,round(ar2,2))
rownames(x) = c(rnames,"sigma","R2","Adjusted R2")
ego1regout = x
####Local-Network Structure
#SUR
clmsurelocal = systemfit(list(clclp,clslp,clvlp),method="SUR",data=
    data_local)
#Format these nice:
rnames = c()
sigmas = c()
r2 = c()
ar2 = c()
x=c()
for(i in clmsurelocal$eq){
  si = summary(i)
  coef = si$coefficients
  rnames = rownames(coef)
  sig = coef[,4]
```

```
  sig = sapply(sig,function(x){ifelse(x<.05,"*","")})
  co = round(coef[,1],2)
  toBind2 = paste("(",round(coef[,2],2),")",sep="")
  toBind = paste(co,toBind2,sig,sep="")
  x=cbind(x,toBind)
  r2=c(r2,si$r.squared)
  ar2 = c(ar2,si$adj.r.squared)
  sigmas = c(sigmas,si$sigma)
}
mnames = c("c","s","v")
colnames(x) = mnames
x = rbind(x,round(sigmas,2))
x = rbind(x,round(r2,2))
x = rbind(x,round(ar2,2))
rownames(x) = c(rnames,"sigma","R2","Adjusted R2")
localregout = x
##########################################
#Full System Linear Regression for Each Level
#(Cuvi-Linear Model)
##########################################
#All Levels in one model
#SUR


model = list(pclace2,pclase2,pclave2,pclsce2,pclsse2,pclsve2,pclcce2,
    pclcse2,pclcve2,pclace1,pclase1,pclave1,pclsce1,pclsse1,pclsve1,
    pclcce1,pclcse1,pclcve1,pclclp,pclslp,pclvlp)
pooled_fit=systemfit(model,method="SUR",data=data_pooled)

#Get Results into a nice table format

##########################################
#Structure Comparisons
#
##########################################
cen = t.test(data_ego1$Centrality,data_ego2$Centrality,alternative="
    less")
```

```
dis = t.test(data_ego1$Distance,data_ego2$Distance,alternative="less")
con = t.test(data_ego1$Connectivity,data_ego2$Connectivity,alternative="less")
fed = t.test(data_ego1$Feedback,data_ego2$Feedback,alternative="less")
clu = t.test(data_ego1$Clustering,data_ego2$Clustering,alternative="less")

cen = t.test(data_local$Centrality,data_ego1$Centrality,alternative="less")
dis = t.test(data_local$Distance,data_ego2$Distance,alternative="less")
con = t.test(data_local$Connectivity,data_ego1$Connectivity,alternative="less")
fed = t.test(data_local$Feedback,data_ego1$Feedback,alternative="less")
clu = t.test(data_local$Clustering,data_ego1$Clustering,alternative="less")
vel = t.test(data_local$velocity,data_ego1$velocity,alternative="less")

cen = t.test(data_node$Centrality,data_edge$Centrality,alternative="less")
dis = t.test(data_node$Distance,data_edge$Distance,alternative="less")
con = t.test(data_node$Connectivity,data_edge$Connectivity,alternative="less")
fed = t.test(data_node$Feedback,data_edge$Feedback,alternative="less")
clu = t.test(data_node$Clustering,data_edge$Clustering,alternative="less")
vel = t.test(data_node$velocity,data_edge$velocity,alternative="less")
```

```
##########################################
#Regression Restrictions
#
##########################################
m=matrix(0,nrow=45,ncol = 144)
 count = 0
 rCount = 1
 for(i in 0:8){
   b=12+16*i
   e=b+4
   for(j in b:e){
    m[rCount,j] = 1
    rCount = rCount+1
   }
 }

 m = matrix(0,nrow = 108, ncol = 117)
 count = 2
 for(i in 1:108){
   m[i,count] = 1
   count=count+1
   if(count%%14==0)
     count = count+1
 }
 for(i in 2:117){
   if()
 }
 clmsurego2
```

Listing B.1: R Listing Data Analysis

References

Abolghasemi, M., Khodakarami, V., & Tehranifard, H. (2015). A new approach for supply chain risk management: Mapping scor into bayesian network. *Journal of industrial engineering and management*, *8*(1), 280–302.

Adebola Adeseun, M., Anosike, A., Garza-Reyes, J. A., & Al-Talib, M. (2018). Supply chain risk perception: understanding the gap between theory and practice. In *Proceedings of the 16th ifac symposium on information control problems in manufacturing (incom).*

Agrogiannis, S., & Kinias, I. (2018). Sustainability dilemmas: Risk and uncertainty in supply chain management. *Journal of Business Theory and Practice*, *6*(2), 95.

Ahuja, G., Soda, G., & Zaheer, A. (2012). The genesis and dynamics of organizational networks. *Organization science*, *23*(2), 434–448.

Anderson, J. C., Håkansson, H., & Johanson, J. (1994). Dyadic business relationships within a business network context. *The Journal of Marketing*, 1–15.

Aqlan, F., & Lam, S. S. (2015a). A fuzzy-based integrated framework for supply chain risk assessment. *International Journal of Production Economics*, *161*, 54–63.

Aqlan, F., & Lam, S. S. (2015b). Supply chain risk modelling and mitigation. *International Journal of Production Research*, *53*(18), 5640–5656.

Aqlan, F., & Lam, S. S. (2016). Supply chain optimization under risk and uncertainty: A case study for high-end server manufacturing. *Computers & Industrial Engineering*, *93*, 78–87.

Arshinder, K., Kanda, A., & Deshmukh, S. (2011). A review on supply chain coordination: coordination mechanisms, managing uncertainty and research directions. In *Supply chain coordination under uncertainty* (pp. 39–82). Springer.

Autry, C. W., & Griffis, S. E. (2008). Supply chain capital: the impact of structural and relational linkages on firm execution and innovation. *Journal of business logistics*,

29(1), 157–173.

Baharmand, H., Comes, T., & Lauras, M. (2017). Managing in-country transportation risks in humanitarian supply chains by logistics service providers: Insights from the 2015 nepal earthquake. *International journal of disaster risk reduction*, *24*, 549–559.

Bakshi, N., & Mohan, S. (2015). Mitigating disruption cascades in supply networks. *Preprint at http://faculty. london. edu/nbakshi/Nw. pdf*.

Barney, J., Wright, M., & Ketchen Jr, D. J. (2001). The resource-based view of the firm: Ten years after 1991. *Journal of management*, *27*(6), 625–641.

Barratt, M., & Oke, A. (2007). Antecedents of supply chain visibility in retail supply chains: a resource-based theory perspective. *Journal of operations management*, *25*(6), 1217–1233.

Barry, J. (2004). Supply chain risk in an uncertain global supply chain environment. *international journal of physical distribution & logistics management*, *34*(9), 695–697.

Basole, R. C., & Bellamy, M. A. (2014a). Supply network structure, visibility, and risk diffusion: A computational approach. *Decision Sciences*, *45*(4), 753–789.

Basole, R. C., & Bellamy, M. A. (2014b). Visual analysis of supply network risks: Insights from the electronics industry. *Decision Support Systems*, *67*, 109–120.

Basole, R. C., & Rouse, W. B. (2008). Complexity of service value networks: Conceptualization and empirical investigation. *IBM systems journal*, *47*(1), 53–70.

Behzadi, G., O'Sullivan, M. J., Olsen, T. L., Scrimgeour, F., & Zhang, A. (2017). Robust and resilient strategies for managing supply disruptions in an agribusiness supply chain. *International Journal of Production Economics*, *191*, 207–220.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations research*, *30*(5), 961–981.

Bellamy, M. A., & Basole, R. C. (2013). Network analysis of supply chain systems: A systematic review and future research. *Systems Engineering*, *16*(2), 235–249.

Bernardes, E. S. (2010). The effect of supply management on aspects of social capital and the impact on performance: A social network perspective. *Journal of supply*

chain management, *46*(1), 45–55.

Bernstein, R. J. (2011). *Beyond objectivism and relativism: Science, hermeneutics, and praxis.* University of Pennsylvania Press.

Blackhurst*, J., Craighead, C. W., Elkins, D., & Handfield, R. B. (2005). An empirically derived agenda of critical research issues for managing supply-chain disruptions. *International journal of production research*, *43*(19), 4067–4081.

Blackhurst, J., Dunn, K. S., & Craighead, C. W. (2011). An empirically derived framework of global supply resiliency. *Journal of Business Logistics*, *32*(4), 374–391.

Blackhurst, J., Rungtusanatham, M. J., Scheibe, K., & Ambulkar, S. (2018). Supply chain vulnerability assessment: A network based visualization and clustering analysis approach. *Journal of Purchasing and Supply Management*, *24*(1), 21–30.

Blome, C., Schoenherr, T., & Rexhausen, D. (2013). Antecedents and enablers of supply chain agility and its effect on performance: a dynamic capabilities perspective. *International Journal of Production Research*, *51*(4), 1295–1318.

Blos, M. F., da Silva, R. M., & Wee, H.-M. (2018). A framework for designing supply chain disruptions management considering productive systems and carrier viewpoints. *International Journal of Production Research*, 1–17.

Blos, M. F., Hoeflich, S. L., & Miyagi, P. E. (2015). A general supply chain continuity management framework. *Procedia Computer Science*, *55*, 1160–1164.

Bode, C., & Wagner, S. M. (2015). Structural drivers of upstream supply chain complexity and the frequency of supply chain disruptions. *Journal of Operations Management*, *36*, 215–228.

Bode, C., Wagner, S. M., Petersen, K. J., & Ellram, L. M. (2011). Understanding responses to supply chain disruptions: Insights from information processing and resource dependence perspectives. *Academy of Management Journal*, *54*(4), 833–856.

Borgatti, S. P., & Everett, M. G. (2006). A graph-theoretic perspective on centrality. *Social networks*, *28*(4), 466–484.

Borgatti, S. P., & Li, X. (2009). On social network analysis in a supply chain context.

Journal of Supply Chain Management, *45*(2), 5–22.

Bozarth, C. C., Warsing, D. P., Flynn, B. B., & Flynn, E. J. (2009). The impact of supply chain complexity on manufacturing plant performance. *Journal of Operations Management*, *27*(1), 78–93.

Braunscheidel, M. J., & Suresh, N. C. (2018). Cultivating supply chain agility: Managerial actions derived from established antecedents. In *Supply chain risk management* (pp. 289–309). Springer.

Brintrup, A., & Ledwoch, A. (2018). Supply network science: Emergence of a new perspective on a classical field. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *28*(3), 033120.

Bromiley, P., & Rau, D. (2014). Towards a practice-based view of strategy. *Strategic Management Journal*, *35*(8), 1249–1256.

Brusset, X., & Teller, C. (2017). Supply chain capabilities, risks, and resilience. *International Journal of Production Economics*, *184*, 59–68.

Bugerta, N., & Lascha, R. (2018). Effectiveness of responsive pricing in the face of supply chain disruptions. *Computers & Industrial Engineering*.

Burgess, K., Singh, P. J., & Koroglu, R. (2006). Supply chain management: a structured literature review and implications for future research. *International Journal of Operations & Production Management*, *26*(7), 703–729.

Burt, R. S. (1982). Toward a structural theory of action: network models of social structure, perception, and action.

Busse, C. (2016). Doing well by doing good? the self-interest of buying firms and sustainable supply chain management. *Journal of Supply Chain Management*, *52*(2), 28–47.

Busse, C., Schleper, M. C., Weilenmann, J., & Wagner, S. M. (2017). Extending the supply chain visibility boundary: utilizing stakeholders for identifying supply chain sustainability risks. *International Journal of Physical Distribution & Logistics Management*, *47*(1), 18–40.

Capaldo, A., & Giannoccaro, I. (2015a). How does trust affect performance in the supply chain? the moderating role of interdependence. *International Journal of*

Production Economics, *166*, 36–49.

Capaldo, A., & Giannoccaro, I. (2015b). Interdependence and network-level trust in supply chain networks: A computational study. *Industrial Marketing Management*, *44*, 180–195.

Carbonara, N., & Pellegrino, R. (2017). Real options approach to evaluate postponement as supply chain disruptions mitigation strategy. *International Journal of Production Research*, 1–23.

Carnovale, S. (2014). *The antecedents of supply chain manufacturing joint ventures: a social network perspective* (Unpublished doctoral dissertation). Rutgers University-Graduate School-Newark.

Carnovale, S., Rogers, D. S., & Yeniyurt, S. (2016). Bridging structural holes in global manufacturing equity based partnerships: A network analysis of domestic vs. international joint venture formations. *Journal of Purchasing and Supply Management*, *22*(1), 7–17.

Carnovale, S., & Yeniyurt, S. (2014). The role of ego networks in manufacturing joint venture formations. *Journal of Supply Chain Management*, *50*(2), 1–17.

Carnovale, S., & Yeniyurt, S. (2015). The impact of supply network structure on the financial performance of the firm. In *Supply chain forum: An international journal* (Vol. 16, pp. 18–28).

Carnovale, S., Yeniyurt, S., & Rogers, D. S. (2017). Network connectedness in vertical and horizontal manufacturing joint venture formations: A power perspective. *Journal of Purchasing and Supply Management*, *23*(2), 67–81.

Carter, C. R., Ellram, L. M., & Tate, W. (2007). The use of social network analysis in logistics research. *Journal of Business Logistics*, *28*(1), 137–168.

Carter, C. R., Kosmol, T., & Kaufmann, L. (2017). Toward a supply chain practice view. *Journal of Supply Chain Management*, *53*(1), 114–122.

Carter, C. R., Leuschner, R., & Rogers, D. S. (2007). A social network analysis of the journal of supply chain management: knowledge generation, knowledge diffusion and thought leadership. *Journal of Supply Chain Management*, *43*(2), 15–28.

Carter, C. R., & Rogers, D. S. (2008). A framework of sustainable supply chain management: moving toward new theory. *International journal of physical distribution & logistics management*, *38*(5), 360–387.

Carter, C. R., Rogers, D. S., & Choi, T. Y. (2015). Toward the theory of the supply chain. *Journal of Supply Chain Management*, *51*(2), 89–97.

Casciaro, T., & Piskorski, M. J. (2005). Power imbalance, mutual dependence, and constraint absorption: A closer look at resource dependence theory. *Administrative science quarterly*, *50*(2), 167–199.

Cavinato, J. L. (2004). Supply chain logistics risks: From the back room to the board room. *International journal of physical distribution & logistics management*, *34*(5), 383–387.

Chaghooshi, A., Momeni, M., Abdollahi, B., Safari, H., & Kamalabadi, I. (2018). Planning for disruptions in supply chain networks. *Uncertain Supply Chain Management*, *6*(2), 135–148.

Chapman, P., Christopher, M., Jüttner, U., Peck, H., & Wilding, R. (2002). Identifying and managing supply chain vulnerability. *Logistics & Transport Focus*, *4*(4), 59–70.

Chen, I. J., & Paulraj, A. (2004). Towards a theory of supply chain management: the constructs and measurements. *Journal of operations management*, *22*(2), 119–150.

Chen, P.-S., & Wu, M.-T. (2013). A modified failure mode and effects analysis method for supplier selection problems in the supply chain risk environment: A case study. *Computers & Industrial Engineering*, *66*(4), 634–642.

Chen, X., Xi, Z., & Jing, P. (2017). A unified framework for evaluating supply chain reliability and resilience. *IEEE Transactions on Reliability*, *66*(4), 1144–1156.

Choi, T. Y., Dooley, K. J., & Rungtusanatham, M. (2001). Supply networks and complex adaptive systems: control versus emergence. *Journal of operations management*, *19*(3), 351–366.

Choi, T. Y., & Hong, Y. (2002). Unveiling the structure of supply networks: case studies in honda, acura, and daimlerchrysler. *Journal of Operations Management*, *20*(5),

469–493.

Choi, T. Y., & Kim, Y. (2008). Structural embeddedness and supplier management: A network perspective. *Journal of Supply Chain Management*, *44*(4), 5–13.

Choi, T. Y., & Krause, D. R. (2006). The supply base and its complexity: Implications for transaction costs, risks, responsiveness, and innovation. *Journal of Operations Management*, *24*(5), 637–652.

Choi, T. Y., & Wu, Z. (2009). Triads in supply networks: Theorizing buyer–supplier–supplier relationships. *Journal of Supply Chain Management*, *45*(1), 8–25.

Chopra, S., & Sodhi, M. (2004). Supply-chain breakdown. *MIT Sloan management review*, *46*(1), 53–61.

Christopher, M., & Peck, H. (2004). Building the resilient supply chain. *The international journal of logistics management*, *15*(2), 1–14.

Churchill Jr, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of marketing research*, 64–73.

Cohen, M. A., & Kunreuther, H. (2007). Operations risk management: Overview of paul kleindorfer's contributions. *Production and Operations Management*, *16*(5), 525–541.

Cooper, M. C., & Ellram, L. M. (1993). Characteristics of supply chain management and the implications for purchasing and logistics strategy. *The international journal of logistics management*, *4*(2), 13–24.

Cooper, M. C., Lambert, D. M., & Pagh, J. D. (1997). Supply chain management: more than a new name for logistics. *The international journal of logistics management*, *8*(1), 1–14.

Craighead, C. W., Blackhurst, J., Rungtusanatham, M. J., & Handfield, R. B. (2007). The severity of supply chain disruptions: design characteristics and mitigation capabilities. *Decision Sciences*, *38*(1), 131–156.

Cristopher, M. (1992). *Logistics.* Chapman & Hall, London.

Croom, S., Romano, P., & Giannakis, M. (2000). Supply chain management: an analytical framework for critical literature review. *European journal of purchasing & supply management*, *6*(1), 67–83.

Daultani, Y., Goswami, M., Vaidya, O. S., & Kumar, S. (2017). Inclusive risk modeling for manufacturing firms: a bayesian network approach. *Journal of Intelligent Manufacturing*, 1–15.

Delfmann, W., & Albers, S. (2000). Supply chain management in the global context..

Denardo, E. V., & Lee, T. Y. (1996). Managing uncertainty in a serial production line. *Operations Research*, *44*(2), 382–392.

Dhanaraj, C., & Parkhe, A. (2006). Orchestrating innovation networks. *Academy of management review*, *31*(3), 659–669.

Diestel, R. (2017). *Graph theory*. Springer Publishing Company, Incorporated.

Dolgui, A., Ivanov, D., & Sokolov, B. (2018). Ripple effect in the supply chain: an analysis and recent literature. *International Journal of Production Research*, *56*(1-2), 414–430.

Domínguez Cañizares, R. (2014). The impact of supply chain structures on performance.

Dong, M. (2001). *Process modeling, performance analysis and configuration simulation in integrated supply chain network design* (Unpublished doctoral dissertation). Virginia Tech.

Dong, M. C., Liu, Z., Yu, Y., & Zheng, J.-H. (2015). Opportunism in distribution networks: The role of network embeddedness and dependence. *Production and Operations Management*, *24*(10), 1657–1670.

Dong, Q., & Cooper, O. (2016). An orders-of-magnitude ahp supply chain risk assessment framework. *International Journal of Production Economics*, *182*, 144–156.

Drees, J. M., & Heugens, P. P. (2013). Synthesizing and extending resource dependence theory: A meta-analysis. *Journal of Management*, *39*(6), 1666–1698.

Dubois, A., & Fredriksson, P. (2008). Cooperating and competing in supply networks: Making sense of a triadic sourcing strategy. *Journal of Purchasing and Supply Management*, *14*(3), 170–179.

DuHadway, S., Carnovale, S., & Hazen, B. (2017). Understanding risk management for intentional supply chain disruptions: Risk detection, risk mitigation, and risk recovery. *Annals of Operations Research*, 1–20.

DuHadway, S., Carnovale, S., & Kannan, V. (2018). Organizational communication and individual behavior: Implications for supply chain risk management. *Journal of Supply Chain Management*.

Ekwall, D., & Lantz, B. (2018). The use of violence in cargo theft–a supply chain disruption case. *Journal of Transportation Security*, *11*(1-2), 3–21.

Ellram, L. M. (1991). Supply-chain management: the industrial organisation perspective. *International Journal of Physical Distribution & Logistics Management*, *21*(1), 13–22.

Ellram, L. M., & Cooper, M. C. (1990). Supply chain management, partnership, and the shipper-third party relationship. *The International Journal of Logistics Management*, *1*(2), 1–10.

Elluru, S., Gupta, H., Kaur, H., & Singh, S. P. (2017). Proactive and reactive models for disaster resilient supply chain. *Annals of Operations Research*, 1–26.

Emerson, R. M. (1962). Power-dependence relations. *American sociological review*, 31–41.

Euler, L. (1736). The solution of a problem relating to the geometry of position.

Fahimnia, B., Tang, C. S., Davarzani, H., & Sarkis, J. (2015). Quantitative models for managing supply chain risks: A review. *European Journal of Operational Research*, *247*(1), 1–15.

Fan, Y., & Stevenson, M. (2018). A review of supply chain risk management: definition, theory, and research agenda. *International Journal of Physical Distribution & Logistics Management*, *48*(3), 205–230.

Fisher, M. L. (1997). What is the right supply chain for your product? *Harvard business review*, 1.

Flynn, B. B., Koufteros, X., & Lu, G. (2016). On theory in supply chain uncertainty and its implications for supply chain integration. *Journal of Supply Chain Management*, *52*(3), 3–27.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, *1*(3), 215–239.

Friedman, M. (1953). The methodology of positive economics.

Fu, S., Zhang, D., Montewka, J., Yan, X., & Zio, E. (2016). Towards a probabilistic model for predicting ship besetting in ice in arctic waters. *Reliability Engineering & System Safety*, *155*, 124–136.

Galaskiewicz, J. (2011). Studying supply chains from a social network perspective. *Journal of Supply Chain Management*, *47*(1), 4–8.

Gaonkar, R., & Viswanadham, N. (2004). A conceptual and analytical framework for the management of risk in supply chains. In *Robotics and automation, 2004. proceedings. icra'04. 2004 ieee international conference on* (Vol. 3, pp. 2699–2704).

Gaonkar, R. S., & Viswanadham, N. (2007). Analytical framework for the management of risk in supply chains. *IEEE Transactions on automation science and engineering*, *4*(2), 265–273.

Garvey, M. D., Carnovale, S., & Yeniyurt, S. (2015). An analytical framework for supply network risk propagation: A bayesian network approach. *European Journal of Operational Research*, *243*(2), 618–627.

Gaudenzi, B., & Borghesi, A. (2006). Managing risks in the supply chain using the ahp method. *The International Journal of Logistics Management*, *17*(1), 114–136.

Ghadge, A., Dani, S., Chester, M., & Kalawsky, R. (2013). A systems approach for modelling supply chain risks. *Supply chain management: an international journal*, *18*(5), 523–538.

Ghadge, A., Dani, S., & Kalawsky, R. (2011). Systems thinking for modeling risk propagation in supply networks. In *Industrial engineering and engineering management (ieem), 2011 ieee international conference on* (pp. 1685–1689).

Giaglis, G. M., Minis, I., Tatarakis, A., & Zeimpekis, V. (2004). Minimizing logistics risk through real-time vehicle routing and mobile technologies: Research to date and future trends. *International Journal of Physical Distribution & Logistics Management*, *34*(9), 749–764.

Giannakis, M., & Croom, S. R. (2004). Toward the development of a supply chain management paradigm: a conceptual framework. *Journal of supply chain management*, *40*(1), 27–37.

Gilchrist, W. (1993). Modelling failure modes and effects analysis. *International Journal*

of Quality & Reliability Management, *10*(5).

Giri, B., & Sarker, B. R. (2017). Improving performance by coordinating a supply chain with third party logistics outsourcing under production disruption. *Computers & Industrial Engineering*, *103*, 168–177.

Giunipero, L. C., & Aly Eltantawy, R. (2004). Securing the upstream supply chain: a risk management approach. *International Journal of Physical Distribution & Logistics Management*, *34*(9), 698–713.

Govindan, K., & Chaudhuri, A. (2016). Interrelationships of risks faced by third party logistics service providers: A dematel based approach. *Transportation Research Part E: Logistics and Transportation Review*, *90*, 177–195.

Govindan, K., & Fattahi, M. (2017). Investigating risk and robustness measures for supply chain network design under demand uncertainty: A case study of glass supply chain. *International Journal of Production Economics*, *183*, 680–699.

Govindan, K., Fattahi, M., & Keyvanshokooh, E. (2017). Supply chain network design under uncertainty: A comprehensive review and future research directions. *European Journal of Operational Research*, *263*(1), 108–141.

Green, M. (2004). Loss/risk management notes: Survey: Executives rank fire, disruptions top threats. *Bests Review*, *105*(5), 105.

Han, J., & Shin, K. (2016). Evaluation mechanism for structural robustness of supply chain considering disruption propagation. *International Journal of Production Research*, *54*(1), 135–151.

Handfield, R. B., Blackhurst, J., Elkins, D., & Craighead, C. W. (2007). A framework for reducing the impact of disruptions to the supply chain: Observations from multiple executives. *Supply Chain Risk Management: Minimizing Disruption in Global Sourcing. Taylor and Francis, Boca Raton, FL*, 29–49.

Harland, C., Brenchley, R., & Walker, H. (2003). Risk in supply networks. *Journal of Purchasing and Supply management*, *9*(2), 51–62.

Harland, C., Zheng, J., Johnsen, T., & Lamming, R. (2004). A conceptual model for researching the creation and operation of supply networks 1. *British Journal of Management*, *15*(1), 1–21.

Harland, C. M., Lamming, R. C., Zheng, J., & Johnsen, T. E. (2001). A taxonomy of supply networks. *Journal of Supply Chain Management*, *37*(3), 21–27.

Heckerman, D. (1998). A tutorial on learning with bayesian networks. In *Learning in graphical models* (pp. 301–354). Springer.

Heckerman, D., Geiger, D., & Chickering, D. M. (1995). Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, *20*(3), 197–243.

Heckmann, I., Comes, T., & Nickel, S. (2015). A critical review on supply chain risk–definition, measure and modeling. *Omega*, *52*, 119–132.

Hendricks, K. B., & Singhal, V. R. (2005a). Association between supply chain glitches and operating performance. *Management science*, *51*(5), 695–711.

Hendricks, K. B., & Singhal, V. R. (2005b). An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm. *Production and Operations management*, *14*(1), 35–52.

Hinterhuber, A. (2002). Value chain orchestration in action and the case of the global agrochemical industry. *Long range planning*, *35*(6), 615–635.

Ho, D. C., Au, K., & Newton, E. (2002). Empirical research on supply chain management: a critical review and recommendations. *International Journal of Production Research*, *40*(17), 4415–4430.

Ho, W., Zheng, T., Yildiz, H., & Talluri, S. (2015). Supply chain risk management: a literature review. *International Journal of Production Research*, *53*(16), 5031–5069.

Hosseini, S., & Barker, K. (2016). A bayesian network model for resilience-based supplier selection. *International Journal of Production Economics*, *180*, 68–87.

Hou, Y., Wang, X., Wu, Y. J., & He, P. (2018). How does the trust affect the topology of supply chain network and its resilience? an agent-based approach. *Transportation Research Part E: Logistics and Transportation Review*, *116*, 229–241.

Houlihan, J. B. (1985). International supply chain management. *International Journal of Physical Distribution & Materials Management*, *15*(1), 22–38.

Huang, H., He, Y., & Li, D. (2018). Pricing and inventory decisions in the food

supply chain with production disruption and controllable deterioration. *Journal of Cleaner Production*, *180*, 280–296.

Hult, G. T. M., Ketchen Jr, D. J., & Slater, S. F. (2004). Information processing, knowledge development, and strategic supply chain performance. *Academy of management journal*, *47*(2), 241–253.

Ireland, R. D., & Webb, J. W. (2007). A multi-theoretic perspective on trust and power in strategic supply chains. *Journal of Operations management*, *25*(2), 482–497.

Ivanov, D. (2017). Simulation-based ripple effect modelling in the supply chain. *International Journal of Production Research*, *55*(7), 2083–2101.

Ivanov, D., Dolgui, A., & Sokolov, B. (2017). Scheduling of recovery actions in the supply chain with resilience analysis considerations. *International Journal of Production Research*, 1–18.

Ivanov, D., Dolgui, A., Sokolov, B., & Ivanova, M. (2017). Literature review on disruption recovery in the supply chain. *International Journal of Production Research*, *55*(20), 6158–6174.

Ivanov, D., Pavlov, A., Pavlov, D., & Sokolov, B. (2017). Minimization of disruption-related return flows in the supply chain. *International Journal of Production Economics*, *183*, 503–513.

Iyengar, D., Rao, S., & Goldsby, T. J. (2012). The power and centrality of the transportation and warehousing sector within the us economy: A longitudinal exploration using social network analysis. *Transportation Journal*, *51*(4), 373–398.

Jabbarzadeh, A., Fahimnia, B., & Sabouhi, F. (2018). Resilient and sustainable supply chain design: sustainability analysis under disruption risks. *International Journal of Production Research*, 1–24.

Jabbarzadeh, A., Haughton, M., & Khosrojerdi, A. (2018). Closed-loop supply chain network design under disruption risks: A robust approach with real world application. *Computers & Industrial Engineering*, *116*, 178–191.

Jia, F., & Rutherford, C. (2010). Mitigation of supply chain relational risk caused by cultural differences between china and the west. *The International Journal of Logistics Management*, *21*(2), 251–270.

Jiang, J., Liu, C., Huang, H., & Zhang, H. (2017). Risk communication and simulation of supply chain. In *Electronics instrumentation & information systems (eiis), 2017 first international conference on* (pp. 1–5).

Jüttner, U. (2005). Supply chain risk management: Understanding the business requirements from a practitioner perspective. *The international journal of logistics management*, *16*(1), 120–141.

Jüttner, U., Peck, H., & Christopher, M. (2003). Supply chain risk management: outlining an agenda for future research. *International Journal of Logistics: Research and Applications*, *6*(4), 197–210.

Kainuma, Y. (2018). Design and evaluation of global supply chain considering disruption risk. In *Supply chain risk management* (pp. 201–212). Springer.

Kamalahmadi, M., & Parast, M. M. (2017). An assessment of supply chain disruption mitigation strategies. *International Journal of Production Economics*, *184*, 210–230.

Ketchen Jr, D. J., & Hult, G. T. M. (2007). Bridging organization theory and supply chain management: The case of best value supply chains. *Journal of operations management*, *25*(2), 573–580.

Keynes, J. N., et al. (1890). The scope and method of political economy. *History of Economic Thought Books*.

Khan, A. S., Babar, A. R., Tareen, O. T., Naeem, K., & Gailani, M. U. A. (2018). Strategic risk management in supply chain-an empirical investigation. *Archives of Business Research*, *6*(5).

Khojasteh-Ghamari, Z., & Irohara, T. (2018). Supply chain risk management: A comprehensive review. In *Supply chain risk management* (pp. 3–22). Springer.

Kilubi, I. (2016). The strategies of supply chain risk management–a synthesis and classification. *International Journal of Logistics Research and Applications*, *19*(6), 604–629.

Kilubi, I., & Haasis, H. (2015). Supply chain risk management enablers-a framework development through systematic review of the literature from 2000 to 2015.

Kim, J.-B. (2015). Social network analysis of a supply network structural investigation

of the south korean automotive industry. In *Ifip international conference on advances in production management systems* (pp. 332–339).

Kim, Y., Chen, Y.-S., & Linderman, K. (2015). Supply network disruption and resilience: A network structural perspective. *Journal of operations Management*, *33*, 43–59.

Kim, Y., Choi, T. Y., Yan, T., & Dooley, K. (2011). Structural investigation of supply networks: A social network analysis approach. *Journal of Operations Management*, *29*(3), 194–211.

Kito, T., Brintrup, A., New, S., & Reed-Tsochas, F. (2014). The structure of the toyota supply network: an empirical analysis.

Kleindorfer, G. B., O'Neill, L., & Ganeshan, R. (1998). Validation in simulation: various positions in the philosophy of science. *Management Science*, *44*(8), 1087–1099.

Kleindorfer, P. R., & Saad, G. H. (2005). Managing disruption risks in supply chains. *Production and operations management*, *14*(1), 53–68.

Klibi, W., & Martel, A. (2012). Scenario-based supply chain network risk modeling. *European Journal of Operational Research*, *223*(3), 644–658.

Kordlouie, H., Sadeghi, L., & Sadeghi, N. (2018). An investigation on the effect of risk management on earnings volatility for shares of banks listed in tehran stock exchange. *Accounting*, *4*(3), 123–128.

Kotabe, M., Martin, X., & Domoto, H. (2003). Gaining from vertical partnerships: knowledge transfer, relationship duration, and supplier performance improvement in the us and japanese automotive industries. *Strategic management journal*, *24*(4), 293–316.

Kubota, Y. (2016, April 19). Japan earthquakes rattle toyotas vulnerable supply chain. *Wall Street Journal*. Retrieved from `https://www.wsj.com/articles/japan-earthquakes-rattle-toyotas-supply-chain-1460986805`

Kubota, Y., Deng, C., & Craymer, L. (2018, July 5). With tariff deadline at hand, businesses brace for the fallout. *Wall Street Journal*. Retrieved from `https://www.wsj.com/articles/u-s-will-shoot-itself-in-the-foot-if-it-pulls-the-tariff-trigger-china-says-1530767200`

Kumar, M., Basu, P., & Avittathur, B. (2018). Pricing and sourcing strategies for competing retailers in supply chains under disruption risk. *European Journal of Operational Research*, *265*(2), 533–543.

Kurniawan, R., Zailani, S. H., Iranmanesh, M., & Rajagopal, P. (2017). The effects of vulnerability mitigation strategies on supply chain effectiveness: risk culture as moderator. *Supply Chain Management: An International Journal*, *22*(1), 1–15.

Kwak, D.-W., Seo, Y.-J., & Mason, R. (2018). Investigating the relationship between supply chain innovation, risk management capabilities and competitive advantage in global supply chains. *International Journal of Operations & Production Management*, *38*(1), 2–21.

Lambert, D. M., & Cooper, M. C. (2000). Issues in supply chain management. *Industrial marketing management*, *29*(1), 65–83.

Lambert, D. M., Cooper, M. C., & Pagh, J. D. (1998). Supply chain management: implementation issues and research opportunities. *The international journal of logistics management*, *9*(2), 1–20.

Lamming, R., Johnsen, T., Zheng, J., & Harland, C. (2000). An initial classification of supply networks. *International Journal of Operations & Production Management*, *20*(6), 675–691.

Landers, P., & Karioka, K. (2016, April 14). Strong earthquake strikes japanese island of kyushu. *Wall Street Journal*. Retrieved from `https://www.wsj.com/articles/strong-quake-strikes-southern-japanese-island-of-kyushu-1460638093`

Lavastre, O., Gunasekaran, A., & Spalanzani, A. (2014). Effect of firm characteristics, supplier relationships and techniques used on supply chain risk management (scrm): an empirical investigation on french industrial firms. *International Journal of Production Research*, *52*(11), 3381–3403.

Ledwoch, A., Yasarcan, H., & Brintrup, A. (2018). The moderating impact of supply network topology on the effectiveness of risk management. *International Journal of Production Economics*, *197*, 13–26.

Levner, E., & Ptuskin, A. (2018). Entropy-based model for the ripple effect: managing environmental risks in supply chains. *International Journal of Production*

Research, *56*(7), 2539–2551.

Li, C., Ren, J., & Wang, H. (2016). A system dynamics simulation model of chemical supply chain transportation risk management systems. *Computers & Chemical Engineering*, *89*, 71–83.

Li, G., Fan, H., Lee, P. K., & Cheng, T. (2015). Joint supply chain risk management: An agency and collaboration perspective. *International Journal of Production Economics*, *164*, 83–94.

Lin, C.-C., & Wang, T.-H. (2011). Build-to-order supply chain network design under supply and demand uncertainties. *Transportation Research Part B: Methodological*, *45*(8), 1162–1176.

Linh, D. T. T., Amer, Y., Lee, S.-H., & Phuc, P. N. K. (2018). Development of a reverse supply chain model for electronic waste incorporating transportation risk. *IPTEK Journal of Proceedings Series*(3), 30–36.

Liu, L., Liu, X., & Liu, G. (2018). The risk management of perishable supply chain based on coloured petri net modeling. *Information processing in agriculture*, *5*(1), 47–59.

Lockamy III, A. (2011). Benchmarking supplier risks using bayesian networks. *Benchmarking: An International Journal*, *18*(3), 409–427.

Lockamy III, A., & McCormack, K. (2010). Analysing risks in supply networks to facilitate outsourcing decisions. *International Journal of Production Research*, *48*(2), 593–611.

Lockamy III, A., & McCormack, K. (2012). Modeling supplier risks using bayesian networks. *Industrial Management & Data Systems*, *112*(2), 313–333.

Lonsdale, C. (1999). Effectively managing vertical supply relationships: a risk management model for outsourcing. *Supply chain management: An international journal*, *4*(4), 176–183.

MacCarthy, B. L., & Jayarathne, P. (2013). Supply network structures in the international clothing industry: differences across retailer types. *International Journal of Operations & Production Management*, *33*(7), 858–886.

Macdonald, J. R., & Corsi, T. M. (2013). Supply chain disruption management: Severe

events, recovery, and performance. *Journal of Business Logistics*, *34*(4), 270–288.

Manuj, I., Esper, T. L., & Stank, T. P. (2014). Supply chain risk management approaches under different conditions of risk. *Journal of Business Logistics*, *35*(3), 241–258.

Manuj, I., & Mentzer, J. T. (2008). Global supply chain risk management. *Journal of business logistics*, *29*(1), 133–155.

Margolis, J. T., Sullivan, K. M., Mason, S. J., & Magagnotti, M. (2018). A multi-objective optimization model for designing resilient supply chain networks. *International Journal of Production Economics*.

Meena, P. L., & Sarmah, S. P. (2016). Supplier selection and demand allocation under supply disruption risks. *The International Journal of Advanced Manufacturing Technology*, *83*(1-4), 265–274.

Mehrjoo, M., & Pasek, Z. J. (2016). Risk assessment for the supply chain of fast fashion apparel industry: a system dynamics framework. *International Journal of Production Research*, *54*(1), 28–48.

Melo, M. T., Nickel, S., & Saldanha-Da-Gama, F. (2009). Facility location and supply chain management–a review. *European journal of operational research*, *196*(2), 401–412.

Meyr, H., & Stadtler, H. (2005). Types of supply chains. In *Supply chain management and advanced planning* (pp. 65–80). Springer.

Miles, R. E., & Snow, C. C. (2007). Organization theory and supply chain management: An evolving research perspective. *Journal of operations management*, *25*(2), 459–463.

Mital, M., Del Giudice, M., & Papa, A. (2018). Comparing supply chain risks for multiple product categories with cognitive mapping and analytic hierarchy process. *Technological Forecasting and Social Change*, *131*, 159–170.

Mizgier, K. J. (2017). Global sensitivity analysis and aggregation of risk in multi-product supply chain networks. *International Journal of Production Research*, *55*(1), 130–144.

Mizgier, K. J., Pasia, J. M., & Talluri, S. (2017). Multiobjective capital allocation for

supplier development under risk. *International Journal of Production Research*, *55*(18), 5243–5258.

Monroe, R. W., Teets, J. M., & Martin, P. R. (2014). Supply chain risk management: an analysis of sources of risk and mitigation strategies. *International Journal of Applied Management Science*, *6*(1), 4–21.

Montresor, A., De Pellegrini, F., & Miorandi, D. (2013). Distributed k-core decomposition. *IEEE Transactions on parallel and distributed systems*, *24*(2), 288–300.

Nagesh, G. (2016, April 22). Gm to shut four plants temporarily after japan quakes. *Wall Street Journal*. Retrieved from `https://www.wsj.com/articles/gm-to-shut-four-plants-temporarily-after-japan-quakes-1461342005`

Nair, A., Blome, C., Choi, T. Y., & Lee, G. (2018). Re-visiting collaborative behavior in supply networks–structural embeddedness and the influence of contextual changes and sanctions. *Journal of Purchasing and Supply Management*, *24*(2), 135–150.

Nair, A., & Vidal, J. M. (2011). Supply network topology and robustness against disruptions–an investigation using multi-agent model. *International Journal of Production Research*, *49*(5), 1391–1404.

Nakatani, J., Tahara, K., Nakajima, K., Daigo, I., Kurishima, H., Kudoh, Y., ... others (2018). A graph theory-based methodology for vulnerability assessment of supply chains using the life cycle inventory database. *Omega*, *75*, 165–181.

Namdar, J., Li, X., Sawhney, R., & Pradhan, N. (2018). Supply chain resilience for single and multiple sourcing in the presence of disruption risks. *International Journal of Production Research*, *56*(6), 2339–2360.

Narasimhan, R., & Talluri, S. (2009). *Perspectives on risk management in supply chains.* Elsevier.

Neiger, D., Rotaru, K., & Churilov, L. (2009). Supply chain risk identification with value-focused process engineering. *Journal of operations management*, *27*(2), 154–168.

Nguyen, T., Tran, T., Huynh, T., Ho, T., Le, A., & Do, T. (2018). Managing risks in the fisheries supply chain using house of risk framework (hor) and interpretive structural modeling (ism). In *Iop conference series: Materials science and*

engineering (Vol. 337, p. 012030).

Nishat Faisal, M., Banwet, D. K., & Shankar, R. (2007). Information risks management in supply chains: an assessment and mitigation framework. *Journal of Enterprise Information Management*, *20*(6), 677–699.

Nooraie, S. V., & Parast, M. M. (2015). A multi-objective approach to supply chain risk management: Integrating visibility with supply and demand risk. *International Journal of Production Economics*, *161*, 192–200.

Nooraie, S. V., & Parast, M. M. (2016). Mitigating supply chain disruptions through the assessment of trade-offs among risks, costs and investments in capabilities. *International Journal of Production Economics*, *171*, 8–21.

Nuss, P., Chen, W.-Q., Ohno, H., & Graedel, T. (2016). Structural investigation of aluminum in the us economy using network analysis. *Environmental science & technology*, *50*(7), 4091–4101.

Nuss, P., Graedel, T., Alonso, E., & Carroll, A. (2016). Mapping supply chain risk by network analysis of product platforms. *Sustainable Materials and Technologies*, *10*, 14–22.

Nyoman Pujawan, I., & Geraldin, L. H. (2009). House of risk: a model for proactive supply chain risk management. *Business Process Management Journal*, *15*(6), 953–967.

Ojha, R., Ghadge, A., Tiwari, M. K., & Bititci, U. S. (2018). Bayesian network modelling for supply chain risk propagation. *International Journal of Production Research*, 1–25.

Olivares Aguila, J., & ElMaraghy, W. (2018). Structural complexity and robustness of supply chain networks based on product architecture. *International Journal of Production Research*, 1–18.

Oliver, R. K., Webber, M. D., et al. (1982). Supply-chain management: logistics catches up with strategy. *Outlook*, *5*(1), 42–47.

Olson, D. L., & Wu, D. (2011). Risk management models for supply chain: a scenario analysis of outsourcing to china. *Supply Chain Management: An International Journal*, *16*(6), 401–408.

Orenstein, P. (2016). How does supply network evolution and its topological structure impact supply chain performance? In *Stochastic models in reliability engineering, life science and operations management (smrlo), 2016 second international symposium on* (pp. 562–569).

Pai, R. R., Kallepalli, V. R., Caudill, R. J., & Zhou, M. (2003). Methods toward supply chain risk analysis. In *Systems, man and cybernetics, 2003. ieee international conference on* (Vol. 5, pp. 4560–4565).

Park, K., Min, H., & Min, S. (2016). Inter-relationship among risk taking propensity, supply chain security practices, and supply chain disruption occurrence. *Journal of Purchasing and Supply Management*, *22*(2), 120–130.

Pathak, S. D., Day, J. M., Nair, A., Sawaya, W. J., & Kristal, M. M. (2007). Complexity and adaptivity in supply networks: Building supply network theory using a complex adaptive systems perspective. *Decision sciences*, *38*(4), 547–580.

Pathak, S. D., Dilts, D. M., & Mahadevan, S. (2009). Investigating population and topological evolution in a complex adaptive supply network. *Journal of Supply Chain Management*, *45*(3), 54–57.

Paul, S. K., Sarker, R., & Essam, D. (2017). A quantitative model for disruption mitigation in a supply chain. *European Journal of Operational Research*, *257*(3), 881–895.

Paulraj, A., & Chen, I. J. (2007). Environmental uncertainty and strategic supply management: a resource dependence perspective and performance implications. *Journal of Supply Chain Management*, *43*(3), 29–42.

Paulsson, U. (2004). Supply chain risk management. *Supply chain risk*, 79–96.

Pearl, J. (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier.

Peck, H. (2005). Drivers of supply chain vulnerability: an integrated framework. *International journal of physical distribution & logistics management*, *35*(4), 210–232.

Pedroso, M. C., & Nakano, D. (2009). Knowledge and information flows in supply chains: A study on pharmaceutical companies. *International journal of production*

economics, *122*(1), 376–384.

Pettit, T. J., Fiksel, J., & Croxton, K. L. (2010). Ensuring supply chain resilience: development of a conceptual framework. *Journal of business logistics*, *31*(1), 1–21.

Pfeffer, J., & Salancik, G. R. (2003). *The external control of organizations: A resource dependence perspective*. Stanford University Press.

Pfohl, H.-C., Gallus, P., & Thomas, D. (2011). Interpretive structural modeling of supply chain risks. *International Journal of physical distribution & logistics management*, *41*(9), 839–859.

Pfohl, H.-C., Köhler, H., & Thomas, D. (2010). State of the art in supply chain risk management research: empirical and conceptual findings and a roadmap for the implementation in practice. *Logistics research*, *2*(1), 33–44.

Polyviou, M., Rungtusanatham, M. J., Reczek, R. W., & Knemeyer, A. M. (2018). Supplier non-retention post disruption: What role does anger play? *Journal of Operations Management*.

Popper, K. R. (1959). *The logic of scientific discovery*. University Press.

Porter, M. E. (1991). Towards a dynamic theory of strategy. *Strategic management journal*, *12*(S2), 95–117.

PrasannaVenkatesan, S., & Goh, M. (2016). Multi-objective supplier selection and order allocation under disruption risk. *Transportation Research Part E: Logistics and Transportation Review*, *95*, 124–142.

Prelec, D., & Loewenstein, G. (1991). Decision making over time and under uncertainty: A common approach. *Management science*, *37*(7), 770–786.

Qazi, A., Dickson, A., Quigley, J., & Gaudenzi, B. (2018). Supply chain risk network management: A bayesian belief network and expected utility based approach for managing supply chain risks. *International Journal of Production Economics*, *196*, 24–42.

Qazi, A., Quigley, J., & Dickson, A. (2015). Supply chain risk management: Systematic literature review and a conceptual framework for capturing interdependencies between risks. In *Industrial engineering and operations management (ieom), 2015*

international conference on (pp. 1–13).

Qazi, A., Quigley, J., & Dickson, A. (2018). Cost-effectiveness and manageability based prioritisation of supply chain risk mitigation strategies. In *Supply chain risk management* (pp. 23–42). Springer.

Qazi, A., Quigley, J., Dickson, A., & Ekici, Ş. Ö. (2017). Exploring dependency based probabilistic supply chain risk measures for prioritising interdependent risks and strategies. *European Journal of Operational Research*, *259*(1), 189–204.

Qazi, A., Quigley, J., Dickson, A., & Gaudenzi, B. (2015). A new modelling approach of evaluating preventive and reactive strategies for mitigating supply chain risks. In *International conference on computational logistics* (pp. 569–585).

Qazi, A., Quigley, J., Dickson, A., Gaudenzi, B., & Ekici, Ş. Ö. (2015). Evaluation of control strategies for managing supply chain risks using bayesian belief networks. In *Industrial engineering and systems management (iesm), 2015 international conference on* (pp. 1146–1154).

Qazi, A., Quigley, J., Dickson, A., Gaudenzi, B., & Önsel, Ş. (2015). Selection of optimal redundancy strategies for a supply network. *Innovations and Strate Logistics and Supply Ch*, 419.

Rangel, D. A., de Oliveira, T. K., & Leite, M. S. A. (2015). Supply chain risk classification: discussion and proposal. *International Journal of Production Research*, *53*(22), 6868–6887.

Rao, S., & Goldsby, T. J. (2009). Supply chain risks: a review and typology. *The International Journal of Logistics Management*, *20*(1), 97–123.

Revilla, E., & Saenz, M. J. (2017). The impact of risk management on the frequency of supply chain disruptions: a configurational approach. *International Journal of Operations & Production Management*, *37*(5), 557–576.

Rezapour, S., Farahani, R. Z., & Pourakbar, M. (2017). Resilient supply chain network design under competition: a case study. *European Journal of Operational Research*, *259*(3), 1017–1035.

Ritchie, B., & Brindley, C. (2007). Supply chain risk management and performance: A guiding framework for future development. *International Journal of Operations*

& Production Management, *27*(3), 303–322.

Rivkin, J. W., & Siggelkow, N. (2007). Patterned interactions in complex systems: Implications for exploration. *Management Science*, *53*(7), 1068–1085.

Ro, Y. K., Su, H.-C., & Chen, Y.-S. (2016). A tale of two perspectives on an impending supply disruption. *Journal of Supply Chain Management*, *52*(1), 3–20.

Rosales, C. R., Whipple, J. M., & Blackhurst, J. (2018). The impact of distribution channel decisions and repeated stockouts on manufacturer and retailer performance. *IEEE Transactions on Engineering Management*(99), 1–13.

Rungtusanatham, M., Salvador, F., Forza, C., & Choi, T. Y. (2003). Supply-chain linkages and operational performance: A resource-based-view perspective. *International Journal of Operations & Production Management*, *23*(9), 1084–1099.

Russell, B. (1912). *The problems of philosophy.* Brighthouse.

Russell, B. (1919). *Introduction to mathematical philosophy george allen and unwin.* London.

Sáenz, M. J., Revilla, E., & Acero, B. (2018). Aligning supply chain design for boosting resilience. *Business Horizons*, *61*(3), 443–452.

Safdar, K., & Minaya, E. (2016, March 17). Aropostale says it is exploring strategic alternatives, in dispute with supplier. *Wall Street Journal*. Retrieved from `https://www.wsj.com/articles/aeropostale-posts-another-loss-explores-strategic-review-1458250252`

Salam, M. A., Ali, M., & Kan, K. A. S. (2017). Analyzing supply chain uncertainty to deliver sustainable operational performance: Symmetrical and asymmetrical modeling approaches. *Sustainability*, *9*(12), 2217.

Samvedi, A., Jain, V., & Chan, F. T. (2013). Quantifying risks in a supply chain through integration of fuzzy ahp and fuzzy topsis. *International Journal of Production Research*, *51*(8), 2433–2442.

Sawik, T. (2016). On the risk-averse optimization of service level in a supply chain under disruption risks. *International Journal of Production Research*, *54*(1), 98–113.

Sawik, T. (2017). A portfolio approach to supply chain disruption management. *International Journal of Production Research*, *55*(7), 1970–1991.

Sawik, T. (2018a). Disruption mitigation and recovery in supply chains using portfolio approach. *Omega*.

Sawik, T. (2018b). Selection of dynamic supply portfolio. In *Supply chain disruption management using stochastic mixed integer programming* (pp. 43–67). Springer.

Scheibe, K. P., & Blackhurst, J. (2018). Supply chain disruption propagation: a systemic risk and normal accident theory perspective. *International Journal of Production Research*, *56*(1-2), 43–59.

Schilling, M. A., & Phelps, C. C. (2007). Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. *Management science*, *53*(7), 1113–1126.

Schmitt, T. G., Kumar, S., Stecke, K. E., Glover, F. W., & Ehlen, M. A. (2017). Mitigating disruptions in a multi-echelon supply chain using adaptive ordering. *Omega*, *68*, 185–198.

Shin, K., Shin, Y., Kwon, J.-H., & Kang, S.-H. (2012). Risk propagation based dynamic transportation route finding mechanism. *Industrial Management & Data Systems*, *112*(1), 102–124.

Singh Srai, J., & Gregory, M. (2008). A supply network configuration perspective on international supply chain development. *International Journal of Operations & Production Management*, *28*(5), 386–411.

Sneath, P. H. (1957a). The application of computers to taxonomy. *Microbiology*, *17*(1), 201–226.

Sneath, P. H. (1957b). Some thoughts on bacterial classification. *Microbiology*, *17*(1), 184–200.

Snyder, L. V., Atan, Z., Peng, P., Rong, Y., Schmitt, A. J., & Sinsoysal, B. (2016). Or/ms models for supply chain disruptions: A review. *IIE Transactions*, *48*(2), 89–109.

Snyder, L. V., Scaparra, M. P., Daskin, M. S., & Church, R. L. (2006). Planning for disruptions in supply chain networks. In *Models, methods, and applications for innovative decision making* (pp. 234–257). INFORMS.

Sokolov, B., Ivanov, D., Dolgui, A., & Pavlov, A. (2016). Structural quantification of the

ripple effect in the supply chain. *International Journal of Production Research*, *54*(1), 152–169.

Sommerfeld, D., Teucke, M., & Freitag, M. (2018). Identification of sensor requirements for a quality data-based risk management in multimodal supply chains. *Procedia CIRP*, *72*, 563–568.

Song, W., Ming, X., & Liu, H.-C. (2017). Identifying critical risk factors of sustainable supply chain management: A rough strength-relation analysis method. *Journal of cleaner production*, *143*, 100–115.

Stevens, G. C. (1989). Integrating the supply chain. *International Journal of Physical Distribution & Materials Management*, *19*(8), 3–8.

Strogatz, S. H. (2001). Exploring complex networks. *nature*, *410*(6825), 268.

Svensson, G. (2002). A typology of vulnerability scenarios towards suppliers and customers in supply chains based upon perceived time and relationship dependencies. *International Journal of Physical Distribution & Logistics Management*, *32*(3), 168–187.

Swaminathan, J. M., Smith, S. F., & Sadeh, N. M. (1998). Modeling supply chain dynamics: A multiagent approach. *Decision sciences*, *29*(3), 607–632.

Takata, S., & Yamanaka, M. (2013). Bom based supply chain risk management. *CIRP Annals-Manufacturing Technology*, *62*(1), 479–482.

Talluri, S., Narasimhan, R., & Nair, A. (2006). Vendor performance with supply risk: A chance-constrained dea approach. *International Journal of Production Economics*, *100*(2), 212–222.

Tang, C. S. (2006). Perspectives in supply chain risk management. *International journal of production economics*, *103*(2), 451–488.

Tang, L., Jing, K., He, J., & Stanley, H. E. (2016). Complex interdependent supply chain networks: Cascading failure and robustness. *Physica A: Statistical Mechanics and its Applications*, *443*, 58–69.

Terpend, R., & Ashenbaum, B. (2012). The intersection of power, trust and supplier network size: Implications for supplier performance. *Journal of Supply Chain Management*, *48*(3), 52–77.

Torres-Ruiz, A., & Ravindran, A. R. (2018). Multiple criteria framework for the sustainability risk assessment of a supplier portfolio. *Journal of Cleaner Production*, *172*, 4478–4493.

Trkman, P., & McCormack, K. (2009). Supply chain risk in turbulent environmentsa conceptual model for managing supply chain network risk. *International Journal of Production Economics*, *119*(2), 247–258.

Truong Quang, H., & Hara, Y. (2018). Risks and performance in supply chain: the push effect. *International Journal of Production Research*, *56*(4), 1369–1388.

Tummala, R., & Schoenherr, T. (2011). Assessing and managing risks using the supply chain risk management process (scrmp). *Supply Chain Management: An International Journal*, *16*(6), 474–483.

Valenzuela, J. F. B., Fu, X., Xiao, G., & Goh, R. S. M. (2018). A network-based impact measure for propagated losses in a supply chain network consisting of resilient components. *Complexity*, *2018*.

Varzandeh, J., Farahbod, K., & Zhu, J. J. (2016). Global logistics and supply chain risk management. *Journal of Business and Behavioral Sciences*, *28*(1), 124.

Vereecke, A., Van Dierdonck, R., & De Meyer, A. (2006). A typology of plants in global manufacturing networks. *Management Science*, *52*(11), 1737–1750.

Vilko, J. P., & Hallikas, J. M. (2012). Risk assessment in multimodal supply chains. *International Journal of Production Economics*, *140*(2), 586–595.

Wagner, S. M., & Bode, C. (2008). An empirical examination of supply chain performance along several dimensions of risk. *Journal of business logistics*, *29*(1), 307–325.

Wagner, S. M., & Bode, C. (2009). Dominant risks and risk management practices in supply chains. In *Supply chain risk* (pp. 271–290). Springer.

Wagner, S. M., Mizgier, K. J., & Papageorgiou, S. (2017). Operational disruptions and business cycles. *International journal of production economics*, *183*, 66–78.

Wagner, S. M., & Neshat, N. (2010). Assessing the vulnerability of supply chains using graph theory. *International Journal of Production Economics*, *126*(1), 121–129.

Wagner, S. M., & Neshat, N. (2012). A comparison of supply chain vulnerability indices

for different categories of firms. *International Journal of Production Research*, *50*(11), 2877–2891.

Wall, R. (2016, April 29). Business news: Airbus takes a stumble — jet maker's profit falls 50% as deliveries lag; spending to produce new models saps cash. *Wall Street Journal*. Retrieved from `https://www.wsj.com/articles/airbus-profit-falls-50-in-first-quarter-1461821681`

Wang, X., Chan, H. K., Yee, R. W., & Diaz-Rainey, I. (2012). A two-stage fuzzy-ahp model for risk assessment of implementing green initiatives in the fashion supply chain. *International Journal of Production Economics*, *135*(2), 595–606.

Wang, Y., & Xiao, R. (2016). An ant colony based resilience approach to cascading failures in cluster supply network. *Physica A: Statistical Mechanics and its Applications*, *462*, 150–166.

Wang, Y., & Zhang, F. (2018). Modeling and analysis of under-load-based cascading failures in supply chain networks. *Nonlinear Dynamics*, *92*(3), 1403–1417.

Wang, Z., Hu, H., & Gong, J. (2018). Simulation based multiple disturbances evaluation in the precast supply chain for improved disturbance prevention. *Journal of Cleaner Production*, *177*, 232–244.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small-worldnetworks. *nature*, *393*(6684), 440.

Wei, H.-L., & Wang, E. T. (2010). The strategic value of supply chain visibility: increasing the ability to reconfigure. *European Journal of Information Systems*, *19*(2), 238–249.

Weng, S.-J., Wu, T., Blackhurst, J., & Mackulak, G. (2009). An extended dea model for hospital performance evaluation and improvement. *Health Services and Outcomes Research Methodology*, *9*(1), 39.

Wieland, A., & Marcus Wallenburg, C. (2012). Dealing with supply chain risks: Linking risk management practices and strategies to performance. *International Journal of Physical Distribution & Logistics Management*, *42*(10), 887–905.

Wiengarten, F., Humphreys, P., Gimenez, C., & McIvor, R. (2016). Risk, risk management practices, and the success of supply chain integration. *International Journal*

of Production Economics, *171*, 361–370.

Wilson, M. C. (2007). The impact of transportation disruptions on supply chain performance. *Transportation Research Part E: Logistics and Transportation Review*, *43*(4), 295–320.

Winter, A., Deniaud, I., Marmier, F., & Caillaud, E. (2018). A risk assessment model for supply chain design. implementation at kuehne+ nagel luxembourg. In *Logistics operations management (gol), 2018 4th international conference on* (pp. 1–8).

Wu, D., & Olson, D. L. (2008). Supply chain risk, simulation, and vendor selection. *International journal of production economics*, *114*(2), 646–655.

Wu, K.-J., Liao, C.-J., Tseng, M.-L., Lim, M. K., Hu, J., & Tan, K. (2017). Toward sustainability: using big data to explore the decisive attributes of supply chain risks and uncertainties. *Journal of Cleaner Production*, *142*, 663–676.

Wu, T., Blackhurst, J., & Chidambaram, V. (2006). A model for inbound supply risk analysis. *Computers in industry*, *57*(4), 350–365.

Wu, T., Blackhurst, J., & Ogrady, P. (2007). Methodology for supply chain disruption analysis. *International journal of production research*, *45*(7), 1665–1682.

Wu, T., Huang, S., Blackhurst, J., Zhang, X., & Wang, S. (2013). Supply chain risk management: An agent-based simulation to study the impact of retail stockouts. *IEEE Transactions on Engineering Management*, *60*(4), 676–686.

Wu, T., Shunk, D., Blackhurst, J., & Appalla, R. (2007). Aidea: A methodology for supplier evaluation and selection in a supplier-based manufacturing environment. *International journal of manufacturing technology and management*, *11*(2), 174–192.

Wu, Z., Choi, T. Y., & Rungtusanatham, M. J. (2010). Supplier–supplier relationships in buyer–supplier–supplier triads: Implications for supplier performance. *Journal of Operations Management*, *28*(2), 115–123.

Xu, H.-Y., Fu, X., Ponnambalam, L., Namatame, A., Yin, X. F., & Goh, R. S. M. (2015). A model to evaluate risk propagation considering effect of dynamic risk information sharing and multi-sourcing in supply chain networks. In *Industrial*

engineering and engineering management (ieem), 2015 ieee international conference on (pp. 1593–1597).

Yan, T., Choi, T. Y., Kim, Y., & Yang, Y. (2015). A theory of the nexus supplier: A critical supplier from a network perspective. *Journal of Supply Chain Management*, *51*(1), 52–66.

Yu, L., Suojapelto, K., Hallikas, J., & Tang, O. (2008). Chinese ict industry from supply chain perspectivea case study of the major chinese ict players. *International Journal of Production Economics*, *115*(2), 374–387.

Zegordi, S. H., & Davarzani, H. (2012). Developing a supply chain disruption analysis model: Application of colored petri-nets. *Expert Systems with Applications*, *39*(2), 2102–2111.

Zhalechian, M., Torabi, S. A., & Mohammadi, M. (2018). Hub-and-spoke network design under operational and disruption risks. *Transportation Research Part E: Logistics and Transportation Review*, *109*, 20–43.

Zhang, Y., & Lam, J. S. L. (2016). Estimating economic losses of industry clusters due to port disruptions. *Transportation Research Part A: Policy and Practice*, *91*, 17–33.

Zhang, Y., Zhao, C., & Pang, B. (2018). Budget allocation in coping with supply chain disruption risks. *International Journal of Production Research*, 1–16.

Zhao, K., Kumar, A., Harrison, T. P., & Yen, J. (2011). Analyzing the resilience of complex supply network topologies against random and targeted disruptions. *IEEE Systems Journal*, *5*(1), 28–39.

Zhao, K., Scheibe, K., Blackhurst, J., & Kumar, A. (2018). Supply chain network robustness against disruptions: Topological analysis, measurement, and optimization. *IEEE Transactions on Engineering Management*.

Zhao, L., Huo, B., Sun, L., & Zhao, X. (2013). The impact of supply chain risk on supply chain integration and company performance: a global investigation. *Supply Chain Management: An International Journal*, *18*(2), 115–131.

Zokaee, S., Jabbarzadeh, A., Fahimnia, B., & Sadjadi, S. J. (2017). Robust supply chain network design: an optimization model with real world application. *Annals*

of Operations Research, *257*(1-2), 15–44.

Zsidisin, G., & Ellram, L. (2003). An agency theory investigation of supply chain risk management. *Journal of Supply Chain Management*, *39*(3), 15–27.

Zsidisin, G. A. (2003). A grounded definition of supply risk. *Journal of Purchasing and Supply Management*, *9*(5-6), 217–224.

Zsidisin, G. A., Panelli, A., & Upton, R. (2000). Purchasing organization involvement in risk assessments, contingency plans, and risk management: an exploratory study. *Supply Chain Management: An International Journal*, *5*(4), 187–198.

Zsidisin, G. A., & Ritchie, B. (2009). Supply chain risk management–developments, issues and challenges. In *Supply chain risk* (pp. 1–12). Springer.

Zsidisin, G. A., & Wagner, S. M. (2010). Do perceptions become reality? the moderating role of supply chain resiliency on disruption occurrence. *Journal of Business Logistics*, *31*(2), 1–20.