Essays on the Role of Network Structure in Operational Performance

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Kedong Chen

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Kevin Linderman, Adviser
William Li, Co-Adviser

June 2019

Acknowledgements

The completion of my dissertation would not have been possible without the support and nurturing of my advisers. I would like to express my deepest appreciation to my adviser Dr. Kevin W. Linderman, for his unparalleled support, tireless mentoring, and unwavering guidance throughout my doctoral studies. Dr. Linderman has been inspiring my research in the field of supply chain and operations management. His scholarly diligence and meticulous attitude has had a profound impact on me. I am blessed to have Dr. Linderman as my adviser. I would also like to extend my deepest gratitude to my co-adviser Dr. William Li, for his relentless support and incessant encouragement in my research work. I am extremely grateful for Dr. Li's tremendous efforts to establish my research collaboration with Zhejiang Cainiao Supply Chain Management Co., Ltd. Dr. Li has also offered invaluable suggestions and guidance in my pursuit of the M.S. degree in Statistics and the Ph.D. degree in Business Administration. It is my greatest fortune to have Dr. Li as my co-adviser.

I would also like to extend my sincere thanks to my external committee member Dr. Ankur Mani, for his invaluable insight into the modeling of supply chain network, his time, his constructive and ingenious suggestions, and his belief in my work and abilities. Dr. Mani's suggestions and comments have considerably improved my dissertation. I would also like to extend my gratitude to my committee chair Dr. Ximin Natalie Huang, for her valuable advice and contributions to my dissertation and my job searching.

In addition, I would like to acknowledge the supports and assistance from my collaborators and friends, many of whom have invaluable contributions towards the dissertation. Specially, I would like to thank Dr. Hung-Chung Su, who has had tremendous tangible contributions towards the first essay of my dissertation. Without Dr. Su's help in theorizing and conceptualizing warehouse network's operations, Essay 1 would not have been possible. I would also like to specially thank Dr. Yuhong Li, whose help in the agent-based simulation in Essay 3 cannot be over-estimated. I also very much appreciate the strong support of Dr. Lijun Zhu from Zhejiang Cainiao Supply Chain Management Co., Ltd. Finally, I would like to thank all the other faculty members and Ph.D. students in the Department of Supply Chain and Operations for many useful discussions, which have helped me better my dissertation in many ways.

Dedication

To my parents Yanping Chen and Ying Zhou

To my beloved Kai Yang

Abstract

Research on supply chain networks is an important emerging field. A network perspective is essential because a supply chain is more of a network of organizations involved in various stages of manufacturing and product distribution, than independent firms or simple linear chains. In today's volatile world of interdependence and connectivity among firms and facilities, supply chain management must go beyond single organizations and embrace a holistic view of entire networks. Managers who fail to take into account firms' or facilities' relationships with respect to the rest of the network may produce biased performance evaluations and ineffective improvement strategies.

In my dissertation, I investigate the effect of network structure on firms' operational performance. The dissertation consists of three inter-related essays. The first essay explores how a warehouse's inventory efficiency is affected by its structural position in the network. The second essay prescribes optimal strategies to invest resilience resources in the supply chain network against supply shocks. The third essay clarifies the learning behavior of a supply network that improves resilience through its suppliers' disruptions. The dissertation takes a multi-method approach by utilizing data analytics, stochastic optimization, agent-based simulation, multi-level analysis, etc. The dissertation is motivated by and grounded in real supply chains. The network data and the operational context are related to world-renowned manufacturing and/or logistics companies. This dissertation is informed by business practice and difficulties. Its prescriptions and implications will, in turn, inform organizations.

Contents

List of Tables

List of Figures

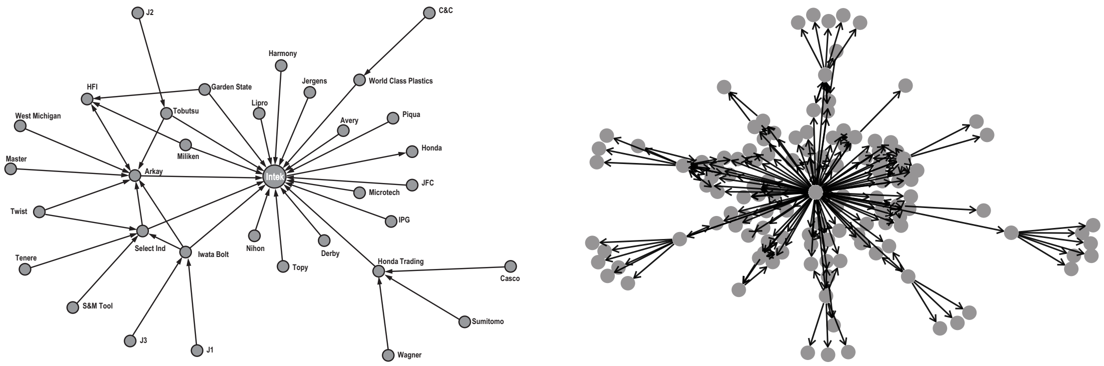
Chapter 1

Introduction

1.1 Problem Statement

In today's world, we are already beyond the point where "supply chain" denotes a linear set of linkages between firms. Rather, firms are faced with a *network* of collaborators, competitors, suppliers, customers, logistic facilities, and intricate inter-organizational relationships. Supply chains now evolve to include a higher level of interdependence and connectivity between more organizations. Business entities are more closely intertwined, and their supply chains essentially become *supply chain networks*. Figure 1.1 displays (a) Acura's supply network and (b) the distribution network of a world leading logistics management company, as the examples of upstream and downstream supply chain networks, respectively.

Research on supply chain networks is emerging. Incorporating the network perspective into supply chain management is gaining increasing recognition. Traditional "linear management" may fail to achieve desired performance improvement due to not considering the effect of network structure on firms' operational performance. For instance, in managing a logistic network, managers who fail to take into account one facility's relationship to the rest of the distribution network may produce biased performance evaluations and ineffective improvement strategies. In fact, many compa-



(a) Materials flow network for Acura, reproduced from Kim et al. (2011)

(b) Sample distribution network of distribution centers and package flows

Figure 1.1: Examples of upstream and downstream supply chain networks

nies, such as Honda, Toyota, BMW, Intel, and Siemens, have realized the importance of "network management" and devoted significant resources to the development of supply network and inter-organizational relationships (Handfield et al., 2010; Liker and Choi, 2004).

This dissertation investigates the role of network structure in the operational performance of both residing organizations and the focal firm. Figure 1.2 presents the overarching framework of the dissertation, where Essay 1 examines how network structure affects individual node's (a warehouse's) performance, Essay 2 proposes optimal resilience investment strategies under supply shocks, through identifying critical nodes in the network, and Essay 3 clarifies the cross-level relationship between node learning and network learning from disruptions. Table 1.1 provides a summary of the three dissertation essays.

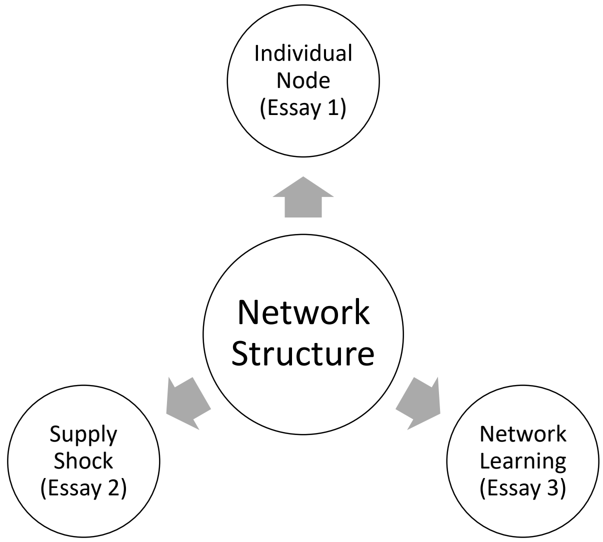


Figure 1.2: Overarching framework of the dissertation

Table 1.1: Summary of the dissertation essays

Essay	Theme	Research question	Method
1	Structural embeddedness and inventory efficiency	How does the structural embeddedness of a warehouse in its network influence the inventory efficiency?	Fixed-effects 2-stage least squares regression model
2	Resilience investment against supply shocks	Where should limited resources be invested in the supply chain network to improve network resilience?	2-stage stochastic program with decision-dependent uncertainty, ordered logistic regression
3	Supply network learning from disruptions	How does the supply network learn from suppliers' disruptions?	Agent-based simulation, multi-level models

1.2 Related Literature Streams

1.2.1 Common Stream – Network Perspective

The three essays build on the same underlying rationale in the network perspective. That is, a network can exert influence, grant opportunities, and create challenges for specific nodes in the network (Borgatti and Li, 2009). A residing organization can experience such structural effect because, when "chain" relationships evolve into a network, the organization's (operational) performance is affected not only by the interplay between its partners and itself, but also more by the interactions among those partners (Choi et al., 2001). Realizing the structural effect in the complicated business world, researchers begin to adopt the network perspective to understand the mechanisms and propose more effective suggestions in the supply chain management.

Starting from basic network elements, Choi and Wu (2009a) proposed the "triad" as the "fundamental building block of a network". Choi and Wu (2009b) exhausted all possible configurations of the buyer-supplier-supplier triad, and explored conditions where buyer-supplier and supplier-supplier relationships emerge, persist, or break down. A series of their related work on triads (e.g., Wu and Choi, 2005; Wu et al., 2010) sheds light on the benefits of examining the network structure in managing even simple business relationships.

As a broader network than triad is investigated, researchers incorporate ideas and concepts from social network analysis, which emphasizes that a company's economic actions are embedded in a network and that their outcomes are substantially influenced by the ongoing pattern of a relationship (Granovetter, 2005; Gulati, 1998). Borgatti and Li (2009) related network concepts specifically to the supply chain context, and suggested that a more systematic adoption of the network perspective would be instrumental in exploring behavioral mechanisms of entire supply networks. Choi et al. (2001) went a step further and understood the supply network as a complex adaptive system, which precisely characterizes the supply network through internal mechanisms, environment, and co-evolution.

In general, the network concepts and measures lay the foundation of this dissertation. Each essay integrates other specific literature streams with network and supply chain management to rationalize the hypotheses and better answer the research questions. The next sub-section describes those related literature streams.

1.2.2 Background Literature for Individual Essays

Background Literature for Essay 1

Essay 1 focuses on *structural embeddedness* that has never been explored in the context of a warehouse network. Structural embeddedness was originally defined as "the impersonal configuration of linkages between people or units" (Nahapiet and Ghoshal, 1998, p. 244). To characterize a warehouse's structural embeddedness, Essay 1 adapts the definition and follows Moran (2005) to distinguish between *direct* and *indirect ties* in a warehouse's ego network. To understand the underlying mechanism of how structural embeddedness affects a warehouse's performance, Essay 1 incorporates the literature of information processing.

Structural embeddedness can affect the amount of information an organization must manage, since ties are "conduits through which resources and information flow" (Borgatti and Foster, 2003, p. 1002). The amount of information is "the volume or quantity of data about organizational activities that is gathered and interpreted by organization participants" (Daft and Macintosh, 1981, p. 210). On the one hand, as the amount of information going through an organization increases, the resources required to effectively process the information also increase. On the other, if alters in an ego network have the flexibility to exchange informational requests among themselves, the focal node can only deal with the needs that cannot be fulfilled by the alters rather than coordinate individually with every alter, which implies a lower total amount of information and reduced information processing needs for the focal node. In general, the integration of information processing with network structure appropriately illustrates the effect of structural embeddedness on a warehouse's inventory efficiency management.

Background Literature for Essay 2

Essay 2 is closely related to the literature of supply chain disruption and resilience management, where scholars have proposed various frameworks and strategies to improve resilience (e.g., Chopra and Sodhi, 2014; Christopher and Peck, 2004; Kleindorfer and Saad, 2005; Sheffi and Rice Jr, 2005; Simchi-Levi et al., 2014, 2015; Tang, 2006). However, much of the emerging literature on supply chain resilience focuses on *how* to reduce disruption and improve resilience. Little is known about *where* to improve resilience in a supply chain network. For instance, Pettit et al. (2013) identified several capabilities that firms need to develop to improve resilience, and argued that a firm's capabilities should align with their disruption vulnerabilities. Snyder and Daskin (2007) found that the design and structure of a supply chain network influences resilience. The average path length (Nair and Vidal, 2011) and structural properties (e.g., clustering coefficient, scale-free and small-world characteristics) (Albert et al., 2000; Basole and Bellamy, 2014; Kim et al., 2015) can significantly influence risk mitigation. Firms should make optimal sourcing decisions based on their positions in the network (Ang et al., 2016; Bimpikis et al., 2017, 2018) and as a response to their suppliers' investments (Bakshi and Mohan, 2015).

Nonetheless, little research, if any, has focused on *where* to invest in resilience in the context of supply chain network. Some studies outside of the supply chain management area have investigated the *where* question. For instance, the network interdiction problem examines which vulnerable nodes to protect to maximize the flow through a network under attacks (e.g., Cormican et al., 1998). This literature stream is critical to managing the infrastructure network such as the highway system (e.g., Peeta et al., 2010). Different from the existing literature, Essay 2 contributes data-driven prescriptions to managing random unintentional disruptions within the context of a supply chain network. It corresponds well with the acknowledgment that improving supply chain resilience should go beyond the focal company's facilities and understand the entire supply chain network (Kim et al., 2015).

Background Literature for Essay 3

Essay 3 examines supply network resilience from a learning perspective. While organizational learning has been extensively studied, the learning behavior of a supply network, in other words, how the network resilience improves due to supplier learning from disruptions, has been largely overlooked. Essay 3 bridges the node-level and network-level learning based on two classic organizational learning curves (Lapré et al., 2011), namely the power curve (e.g., Argote and Epple, 1990; Darr et al., 1995; Dutton and Thomas, 1984; Yelle, 1979) and the exponential curve (e.g., Lapré et al., 2000; Levy, 1965). Scholars have used the power curve to study risk mitigation in airline companies (Haunschild and Sullivan, 2002), US railroads (Baum and Dahlin, 2007), coal mining organizations (Madsen, 2009). The exponential curve has also been applied to other metrics, such as the success rate, the recovery rate, and the production rate (Levy, 1965; Madsen and Desai, 2010; Norrman and Jansson, 2004).

Previous studies in the learning literature have examined multi-level effects, e.g., the effects of population-level (network-level) factors on firm-level performance. For example, Bellamy et al. (2014) studied the influence of supply network structure on firm innovation. Madsen and Desai (2018) investigated the impact of population-level actors on firm failure prevention. However, the learning effect on the network-level performance metrics has not been examined. In practice, firms invest in their suppliers with an overarching aim of improving supply network resilience (Liker and Choi, 2004). Investing in network-level learning may well explain the sustained success of firms like Cisco and Toyota in the face of increasing levels of supplier risks.

1.3 Design of the Dissertation Research

This section introduces the design and results of the three dissertation essays. Brief overarching summary of each essay is given at the beginning of the following subsections.

1.3.1 Research Design of Essay 1 (Chapter 2)

Essay 1, "Structural Embeddedness and Warehouse Inventory Efficiency: A Network Perspective", is developed against the background of emerging omni-channel and e-commerce, where more firms are rethinking and expanding their warehouse networks to better fulfill demand and win over customers. While the managing company sets common requirement of fill rate for the warehouses, inventory efficiency measured as inventory turnover becomes a distinguishing performance metric for each warehouse. Due to that a warehouse's operations may be impacted by other warehouses' actions (i.e., the network structural effect), a warehouse's ability to maintain a high inventory efficiency may depend, in part, on its position in the network. Failing to account for the warehouse's network position can lead to misguided performance evaluation. Essay 1 thus draws on the concept of *structural embeddedness* and contributes to understanding how the direct and indirect ties that constitute structural embeddedness affect warehouse inventory efficiency, which is largely overlooked in the literature.

Essay 1 adopts the fixed-effects 2-stage least squares regression model and examines a proprietary dataset provided by a world leading logistics management company. The empirical analysis controls the warehouse company's inventory policy and sales that may confound the results (Gaur et al., 2005). The analysis uses *inventory turnover* as a measure of inventory efficiency (Lee, 2004), the dependent variable. The independent variables include a warehouse's structural embeddedness, the demand variability and product variety facing the warehouse, and control variables including transshipment and demand. Specifically for structural embeddedness, direct ties are operationalized as the total number of an ego's incoming and outgoing ties (i.e. *degree centrality*) while indirect ties are operationalized as the *proportional density* following past studies (Lokam, 2003; Moran, 2005; Pudlák and Rödl, 1994), which is the ratio between the actual number and all possible ties among alters.

Study results reveal that a warehouse's inventory efficiency significantly depends on its structural embeddedness in the network. Direct ties reduce while indirect ties improve warehouse inventory efficiency, as more direct ties increase the burden of managing information for the ego warehouse, while more indirect ties lessen the information processing burden. Structural embeddedness can significantly interact with product variety and demand variability. The extent of direct ties weakens the negative effect of product variety, while strengthens the negative effect of demand variability on inventory efficiency. In sum, Essay 1 finds that a warehouse's inventory efficiency is a combinative result of its number of direct connections with neighboring warehouses, the variety of products stored within it, and the demand variability faced by the warehouse.

1.3.2 Research Design of Essay 2 (Chapter 3)

Essay 2, "Where to Improve Resilience in the Supply Chain Network under Stochastic Disruptions?", is concerned with today's volatile world full of risks and disruptions such as natural disasters, cyber attacks, epidemics, and political upheavals. These risks and disruptions lead to *supply shocks* that reduce facilities' capacities and decrease the final output of the overall supply chain network. Facing such supply shocks, supply chain managers are highly motivated to invest in resilience resources to mitigate disruptions affecting supply chain actors. However, resource limitations compel managers to identify "critical" facilities whose resilience investment mostly enhances the entire network's output.

Through analyzing the NP-hard 2-stage stochastic program with decision-dependent uncertainty, Essay 2 finds that the optimal investment decision depends on the probability of a disruption and on whether the material flow in the supply chain network can be rerouted. Under rare or frequent disruptions, it is optimal to adopt a node-investment or path-investment strategy, respectively. For mid-level disruptions, the greedy algorithm exhibits comparative performance to the optimal decision of investing on nodes. Through numerical analysis on real distribution networks, Essay 2 generates data-driven prescriptions that are applicable to directed acyclic networks and under various disruption conditions. The commonly acknowledged notion of "node criticality" (Craighead et al., 2007) should be understood in the context of the overall network structure, including in particular *paths*. Supply chain managers can benefit from the proposed algorithm, the path perspective, and the contingency found in Essay 2 – they can sense the weight (importance) of the actors and paths (e.g., shipping routes or product lines) and shift the investment focus accordingly.

1.3.3 Research Design of Essay 3 (Chapter 4)

Essay 3, "Supply Network Resilience Learning: A Data-Driven Exploratory Study", examines how suppliers' learning from disruptions affects the overall supply network's learning to improve resilience. When suppliers experience a disruption, they seek to *learn* from the event and reduce the risk of future events (Fiksel et al., 2015). *Supply network resilience learning* occurs when the supply network becomes more resilient due to individual *supplier learning* from disruptions. Although organizational learning has been extensively studied, supply network learning and its relationship to supplier learning remains unexplored. Essay 3 contributes to clarifying how individual supplier learning translates into supply network learning to better improve network resilience.

Supplier learning at the node level follows the classic power curve and exponential curve to reduce the probability of disruption and increase the probability of recovery, respectively. Through an agent-based simulation model, where suppliers act as agents within an ego network (i.e., the supply base) of a focal firm, Essay 3 examines the curve of network resilience improvement under various disruption probabilities and risk propagation rates. The data analytics uses Honda's and Toyota's supply networks and shows that suppliers' learning-to-prevent improves supply network resilience learning more when suppliers face a lower diffusion rate of a disruption, while learning-to-recover enhances network learning more when the risk of disruption is lower. Essay 3 further shows that more centrally located suppliers influence network learning more, but this depends on the disruption and diffusion of operational risks across the network. Results from Essay 3 suggest practical strategies aimed to make the most of limited resources and facilitate supply network learning to improve the network resilience.

Chapter 2

Structural Embeddedness and Warehouse Inventory Efficiency: A Network Perspective

2.1 Introduction

Retailers are expanding their warehouse networks as part of their strategies to better respond to customers' demands. For instance, Target and Amazon have been building warehouses in the U.S. to satisfy the growing demands from both traditional and e-commerce channels (Stampler, 2019). Similarly, Cainiao Network Inc., the logistics arm of Alibaba, has also been rapidly expanding their distribution network across China, to better meet customer demands (Chou, 2018). As a result, overall warehouse networks become more complex with an increasing number of warehouses and connections between warehouses. For a retailer to maintain a certain service level across warehouses, the efficiency of inventory performance becomes a critical metric to those individual warehouses (Ecklund, 2010). The extent of inventory efficiency relies on individual warehouse's managerial efforts. However, we have limited understanding of how network structural characteristics could influence an individual warehouse's inventory efficiency performance.

Embedded in a warehouse network, a warehouse typically faces a dual challenge of managing warehouse inventories efficiently with certain customer service level target. Anecdotal evidences support the notion that warehouse managers interact with neighboring warehouses for inventory and logistics decisions. In a field study by one of the authors with a globally leading supply chain management company, we observe that, managers in the neighboring warehouses exchange and process inventory and logistics information and make relevant decisions accordingly to manage warehouse inventory. In a sense, coordination in the form of information processing and sharing between neighboring warehouses plays an important role in affecting warehouse inventory efficiency. Zhou and Wan (2017b) found similar mechanisms in their study of a sourcing network, in which individual warehouses coordinate inventory and logistics decisions with each other. Since coordination between neighboring warehouses is necessary for managing inventory, an individual warehouse's ego-network structure essentially represents the inventory coordination network. However, past literature is silent about the role that network structure plays in managing warehouse inventory efficiency.

We posit that a focal warehouse's *structural embeddedness* in the warehouse network can exert a prominent impact on the warehouse's inventory efficiency. Structural embeddedness, in the context of a warehouse network, is defined as the configuration of ties between warehouses (Nahapiet and Ghoshal, 1998). Structural embeddedness has received extensive attention in economics and social science (e.g., Nahapiet and Ghoshal, 1998; Uzzi, 1996; Moran, 2005) – it influences exchanges between entities and further affects entities' behaviors and performances (Granovetter, 1985). Following Moran (2005), we assess a warehouse's structural embeddedness as the warehouse's extent of *direct ties* with neighboring warehouses and *indirect ties* among the neighboring warehouses.

In addition to warehouse network structure, past studies point out that environmental uncertainties could affect warehouse inventory efficiency significantly (e.g., Fisher et al., 1994; Gaur et al., 2005; Smith and Agrawal, 2000; Zipkin, 2000). We learn from the field study that, echoing the previous literature, two kinds of environmental uncertainties can greatly reduce warehouse inventory efficiency: *demand variability* and *product variety*. Demand variability refers to the extent of product demand volatility that each warehouse must fulfill. Product variety relates to the dispersion of the variety of products assigned to each warehouse. These two types of uncertainties reflect the trend that customer demand is difficult to predict; customers prefer a large set of choices (Fisher et al., 1994; Smith and Agrawal, 2000) and firms are increasing their product offerings to increase sales (Smith and Agrawal, 2000). Both demand variability and product variety also exist ubiquitously in other companies (e.g., Zhou and Wan, 2017b) and pose a significant challenge to warehouse inventory management.

Knowing that individual warehouses are embedded in a warehouse network, facing differing extent of product variety and demand variability and striving for inventory efficiency, we ask the following research question: *What is the role of warehouse structural embeddedness, product variety, and demand variability in determining warehouse inventory efficiency?* While there are abundant studies and findings on product variety, demand uncertainty, and inventory management separately (Baker et al. (2017) illustrates related topics thoroughly), very few, if any, have empirically examined warehouse inventory efficiency from a network perspective considering the effects of uncertainties.

We examine the effect of structural embeddedness and its interactions with product variety and demand variability through a proprietary dataset provided by the globally leading supply chain management company (hereafter "Company A"). We specifically consider and address the network endogeneity issue (Carpenter et al., 2012) through a two-stage least-square fixed-effects model with lagged network-related variables as instrumental variables. Study results reveal a negative effect of direct ties and a positive effect of indirect ties on a warehouse's inventory efficiency. More importantly, we find that the extent of direct ties weakens the negative effect of product variety, while strengthens the negative effect of demand variability on inventory efficiency. In sum, we find that a warehouse's inventory efficiency is a combinative result of its number of direct connections with neighboring warehouses, the variety of products stored within it, and the demand variability faced by the warehouse.

We contribute to the operations management literature in the following two aspects. First, we examine the warehouse inventory efficiency from a network perspective. Through warehouse-level panel data, we identify a significant effect of structural embeddedness on warehouse inventory efficiency. Prior studies on inventory performance focus either on firm-level manufacturing practices such as lean and JIT (e.g., Demeter and Matyusz, 2011; Rabinovich et al., 2003), or firm-level factors such as gross margin (e.g., Gaur et al., 2005) – they generally do not consider the effects of network structure. We apply network analysis in a logistics and operational context and extend the network perspective into the field of operations management. Second, we examine interactions between a warehouse's direct ties and the uncertainties faced by a warehouse. These findings provide a better understanding of the role of structural embeddedness, product variety, and demand variability on warehouse inventory efficiency.

Our findings also provide managerial implications regarding both assessment and improvement of warehouse inventory efficiency. In terms of assessment, our results show that warehouse inventory efficiency is determined, in part, by a warehouse's embedded network structure. As firms often rely on auditing (Ackerman, 1990) and benchmarking (Staudt et al., 2015) to compare warehouse inventory performance, our study shows that firms should benchmark individual warehouse efficiency against that of structurally equivalent counterparts. As to performance implications, we help managers better understand the role of a warehouse's network position in managing uncertainties.

The rest of Essay 1 is organized as follows. We review the relevant literature in Section 2.2. Section 2.3 develops the hypotheses. In Section 2.4, we describe the context of warehouse network, the dataset, the variables and measures. Section 2.5 describes the models and results. Section 2.6 discusses the theoretical implications from the analysis, along with the managerial implications. We conclude Essay 1 with future research suggestions in Section 2.7.

2.2 Literature Review

2.2.1 Structural Embeddedness

Nahapiet and Ghoshal (1998, p.244) defined structural embeddedness as "the impersonal config- uration of linkages between people or units", which, in the context of a warehouse network, would be the configuration of ties between networked warehouses. We assess a warehouse's structural embeddedness based on its ego network (Borgatti and Li, 2009). An ego network consists of (1) a focal node (the ego), (2) the set of nodes directly connected to the ego (the alters), and (3) all ties among the ego and alters. Correspondingly, a warehouse's ego network consists of the immediate neighboring warehouses as alters and all the connections among the ego and alters. In a network of warehouses, each warehouse can be the ego. We follow Moran (2005) and characterize structural embeddedness of a warehouse in its ego network in the following two dimensions: (1) *direct ties*: ties between the ego and alters, and (2) *indirect ties*: ties among the alters.

Previous literature supports the notion that structural embeddedness can significantly affect an entity's business performance. In the management settings, Moran

(2005) discovered the importance of structural embeddedness for more routine, execution-oriented tasks (e.g., managerial sales performance). Lin et al. (2009), in the context of government-sponsored R&D consortia, supported significant impacts of structural embeddedness on technology transfer performance. In the supply network context, Choi and Kim (2008) demonstrated the benefits of adopting the structural embeddedness perspective when firms make strategic sourcing decisions. Similarly, Kim (2014) found a positive effect of understanding suppliers' structural embeddedness on a buyer's operational performance. Lawson et al. (2008) also found that structural embeddedness facilitates a buyer's performance improvement. Focusing on distribution networks in China, Dong et al. (2015) showed a distributer's network position affects its opportunism towards supplier through survey data. A recent study by Kao et al. (2017) demonstrated the effects of different network measures on firm-level productivity. Nonetheless, few studies to date have connected structural embeddedness and inventory performance of a warehouse.

2.2.2 Product Variety and Demand Variability

In this section, we review the commonly adopted notion of product variety, i.e., the variety of product SKUs, in the marketing and operations literature (see Ho and Tang (1998) and Ramdas (2003) for a thorough review). Product variety is mostly regarded as a strategy and decision to improve sales or revenues (e.g., Baumol and Ide, 1956; Berry and Cooper, 1999; Smith and Agrawal, 2000). As customers demand more variety of products nowadays (Fisher et al., 1994), firms are increasingly adopting a high variety strategy to increase sales (Smith and Agrawal, 2000).

The operations management literature has pointed out that increase in product variety is associated with a decrease in inventory turnover (Zipkin, 2000). Increasing product variety reduces demand forecast accuracy, hence leading to mismatches between product supply and demand, and ultimately reduced inventory performance

(Fisher and Ittner, 1999; Ton and Raman, 2010; Wan et al., 2012). Wan et al. (2012) found negative but diminishing impacts on product fill rate (as an operational performance measure) as product variety increases, due to increased similarities among products that contribute to demand prediction accuracy. Specifically, for warehouse operations and transshipment, product variety can increase distribution costs as warehouses have to deliver each variety in smaller batches and keep extra inventory to account for unexpected customer demand across a variety of products (Zhou and Wan, 2017a).

Demand variability is the main reason for carrying inventory (Nahmias, 2008). Demand variability can reduce inventory performance, mainly through the safety stock requirement (Zipkin, 2000). Rumyantsev and Netessine (2007) showed that high demand uncertainty leads to high inventory levels for U.S. public firms using a panel dataset from 1992-2002. Further, demand variability has been viewed as the main contributor of the bullwhip effect, which creates difficulty in inventory management across the supply chain and reduces inventory efficiency (Cachon, 1999; Lee et al., 1997b). A recent empirical study by Hançerlioğulları et al. (2016), using the inaccuracy of quarterly sales forecasts as a measure of demand variability, found that demand variability is negatively correlated with firm-level inventory turnover. In the context of a warehouse network, a focal warehouse's demand variability creates uncertainty when the focal warehouse tries to balance the demand and supply (inventory) through sending and receiving products to and from other warehouses.

2.3 Hypothesis Development

A warehouse's structural embeddedness could affect the amount of information it must manage, since network ties are "conduits through which resources and information flow" (Borgatti and Foster, 2003, p.1002). The amount of information is "the volume or quantity of data about organizational activities that is gathered and interpreted by organization participants" (Daft and Macintosh, 1981, p.210). In a network of warehouses, each warehouse needs to exchange operational information (e.g., product, schedule, inventory information) with neighboring warehouses and make logistics decisions (e.g., shipping products, adjusting inventory levels). A warehouse also needs to respond to neighbors' informational requests regarding specific products, inquiries, or shipments. For a warehouse to manage informational exchanges or requests and to coordinate effectively, considerable internal resources such as employees' time and efforts are required. As the amount of information going through a warehouse increases, the required "burden" to effectively process the information also increases.

Considering an ego network, having more direct ties implies that the ego warehouse is connected to more alter warehouses and hence would experience a higher extent of information exchanges and processing due to the sheer number of the alters. A warehouse with more direct ties has lower inventory efficiency because (1) more direct ties lead to more information processing needs and workload, which creates more opportunities of decision errors regarding demand forecast or inventory management (Levinthal, 1997), and (2) the ego warehouse's accommodation of requests from alters could disrupt the existing operations and increase transaction costs (e.g., changing the product mix, searching for a shipment, answering the inquiries) (Choi and Krause, 2006). Therefore, we posit that more direct ties could negatively affect the focal warehouse's inventory efficiency:

Hypothesis 1 *Ceteris paribus*, warehouse inventory efficiency decreases with the extent of direct ties.

As another dimension of structural embeddedness, the extent of indirect ties reflects the interconnectedness in a warehouse's ego network (Podolny and Baron, 1997)

– a higher extent of indirect ties indicates more connections among the alters. Given a specific number of alters, a higher level of indirect ties implies that the alters could exchange and process information among themselves. They could address requests and fulfill each other's orders without involving or disrupting the ego warehouse. In contrast, an ego warehouse with low level of indirect ties may have to respond to every request from the alters (e.g., gathering information, shipping products, making necessary inventory adjustments).

In a sense, with high extent of indirect ties, the ego warehouse can view the connected alters as an entirety and only deal with the needs that cannot be fulfilled by the connected alters rather than coordinate individually with each alter, which implies a lower amount of information processing burden on the ego warehouse. Therefore, the interconnectedness among the alters could benefit the ego warehouse's inventory management, through reducing its workload and errors and causing fewer disruptions to its operations. We thus propose:

Hypothesis 2 *Ceteris paribus*, warehouse inventory efficiency increases with the extent of indirect ties.

We posit that a focal warehouse's extent of direct ties interacts with its product variety. High product variety implies large numbers of different product SKUs and a diverse SKU dispersion that the focal warehouse can allocate to alters to fulfill customer demands. Viewing from an inventory flow perspective (Schmenner and Swink, 1998), having more neighboring warehouses (i.e., more direct ties) essentially enables an ego warehouse with high product variety to ship its products to more outlets, which helps manage the efficiency of the inventory flow for the ego warehouse (Schmenner, 2001; Schmenner and Swink, 1998). In other words, as warehouses have limited capacities in reality, the focal warehouse with only a few direct ties has a lower capability in managing high product variety than a warehouse with more direct ties. A high extent of direct ties prevents the focal warehouse with high product variety to become an inventory bottleneck. Therefore, we propose the following hypothesis:

Hypothesis 3 *Ceteris paribus*, the extent of direct ties weakens the negative effect of product variety on warehouse inventory efficiency.

Past research in operations management has shown that high demand variability increases difficulty in forecasting and inventory management (Gaur et al., 2007). A warehouse facing high demand variability would face challenges in adjusting inventory to match supply and demand, which leads to high inventory level and low inventory turnover (Hançerlioğulları et al., 2016).

We argue that the extent of direct ties also interacts with demand variability. For a focal warehouse with many direct ties, high demand variability increases the information processing and coordination needs between the ego and neighboring warehouses. The coordination needs increase in a nonlinear fashion when an ego warehouse has to coordinate the inventory logistics with an increasing number of neighboring warehouses and an increasing extent of uncertainty (Ladyman et al., 2013). For example, a number of alters could send several requests to the ego warehouse regarding certain products, under high demand variability. The ego would be hesitating or uncertain (acting on the conservative side by holding more inventories) in managing inventory to fulfill alters' demands, which likely reduces inventory efficiency.

Further, under high demand variability, the possibility of decision errors (sending more or less than needed products to neighboring warehouses) also increases, which also reduces inventory efficiency; the focal warehouse may make a wrong decision and ship a product to its neighboring warehouses that is later needed by the focal warehouse. With increasing extent of direct ties and demand variability, managing inventory efficiency becomes an even more difficult task for the focal warehouse. On the contrary, with fewer direct ties and lower demand variability, the focal warehouse can better predict the demand of its own inventory and manage fewer coordination requests from the neighboring warehouses, which helps increase inventory efficiency. Hence, based on the above reasoning, we hypothesize the following:

Hypothesis 4 *Ceteris paribus*, the extent of direct ties strengthens the negative effect of demand variability on warehouse inventory efficiency.

2.4 Research Methods

2.4.1 Context of Warehouse Network

We examine the warehouse network and operations in Company A, a globally leading supply chain management company. Company A provides both service (e.g., cloud-based warehouse management system) and facilities (e.g., warehouses) to worldwide online business-to-consumer (B2C) merchants. Merchants can store, sort, and package products in Company A's warehouses and collaborate with Company A on services such as sales planning, demand forecasting, inventory replenishment, and order delivery. Company A's warehouses store and sort different kinds of products (e.g., television, grocery) and vary in sizes and personnel, but share the same infrastructure, i.e., warehouse management system. Individual warehouses have the discretion to make logistics decisions and arrange inventory transshipment in the warehouse network to balance the yet-to-be realized demand and supply.

We focus on the warehouses of electric appliances as they provide a good research context. Focusing on a single product type reduces the confounding effects of factors like product categories and characteristics. We are able to track the ware-

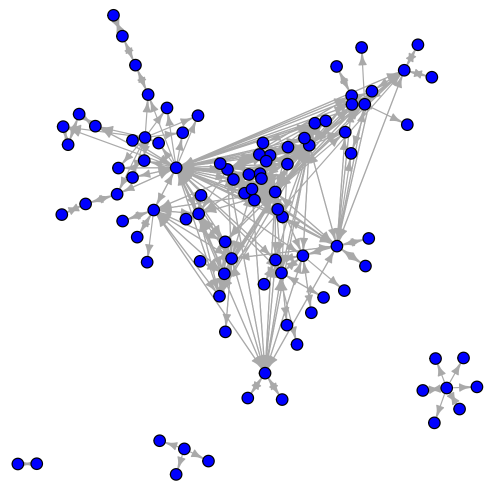


Figure 2.1: A sample warehouse network of electric appliances in week 29

house network within 80 weeks in 2017-2018 (no specific months or dates revealed due to confidentiality). Figure 2.1 displays a sample warehouse network snapshot in week 29. Nodes are warehouses and arrows represent the material flows, which can vary frequently across time. The double arrow represents a mutual flow, indicating bi-directional product flows. Three isolated sub-networks do not have exchanging relationships with other network components in week 29.

2.4.2 Data, Variables, and Measurements

Company A records daily inventory activities of every warehouse. The raw data from the database is at the SKU-supplier-warehouse-day level. We screen out data entries related to electric appliances and aggregate SKUs and suppliers to the warehouse level. The unit of analysis is each warehouse. The aggregation also makes practical sense because once inventory gets into the warehouses, it is managed indifferently regardless of its suppliers. The aggregation can also smooth out random variations among SKUs and suppliers that are unable to be tracked in the original dataset. We further aggregate the daily data into weekly to obtain variables of variability and smooth out daily noises. The final dataset is a panel dataset containing $6,046$ entries at the warehouse-week level with 179 warehouses and 80 weeks.

Warehouse Inventory Efficiency.

We use a commonly used metric – *inventory turnover* – to measure warehouse inventory efficiency (Lee, 2004; Mapes, 2015). Inventory turnover reflects the degree of economic benefits for a warehouse. It is calculated as the total demand quantity over the average inventory in a week. We use the quantity value instead of sales due to the confidential price and cost information. We estimate inventory turnover in the following manner: (1) Calculate each SKU's inventory turnover in a week, and (2) average across all SKUs to obtain the measure. The focus on the electric appliances also helps reduce the product turnover differences.

Structural Embeddedness.

A warehouse's structural embeddedness is characterized as *direct* and *indirect ties* (Moran, 2005). Direct ties are operationalized as the total number of an ego's incoming and outgoing ties (i.e., degree centrality). Each tie represents a link for informational exchanges between the connected warehouses. We consider tie direction because different directions imply different information processing and coordination tasks between warehouses. Indirect ties are operationalized as the proportional density, which is the ratio of the number of ties among alters to the squared number of alters. Aligning with Moran (2005)'s suggestion, the extent of indirect ties need to be scaled to avoid the confounding effect from direct ties (a simple count of indirect ties is highly correlated with or affected by the number of direct ties).

Transshipment.

Transshipment is measured as the total count of transshipment through the ego in the week to and from neighboring warehouses. A warehouse's transshipment can directly affect inventory turnover since it represents the extent of physical product flow in and out of a warehouse. A warehouse with high transshipment essentially implies a high inventory turnover rate, i.e., products flow through the warehouse frequently. We include transshipment to control the physical product flow while structural embeddedness captures the information processing between warehouses for inventory-related decisions.

Network Endogeneity and Instrumental Variables.

The notion that locations of warehouses are not randomly chosen is well established in the empirical literature. For example, Holmes (2011) explicitly discussed the endogeneity issue. Houde et al. (2017) also provided similar evidence and used an instrumental variable (IV) approach to deal with the endogeneity issue. The choice of warehouse locations, hence positions, may depend on proximity to potential customers, competition in the service area, customer characteristics in the service area, etc. Extant literature has indicated that not accounting for endogeneity is a major issue in prior research on networks (Carpenter et al., 2012). To address endogeneity of network-related variables, following previous literature (e.g., Zhou and Wan, 2017b), we use lagged variables (lag by one period) as instrumental variables for transshipment, direct ties and indirect ties.

Lagged variables meet both requirements for instruments – relevance (i.e., identifiability) and exogeneity (i.e., the exclusion restriction). Lagged transshipment, direct and indirect ties are strongly correlated with current transshipment, direct and indirect ties, respectively, and satisfy the identifiability due to the equal number to endogenous variables. In terms of the exclusion restriction, lagged warehouse network-related variables (in the last period) do not directly affect the current period's warehouse inventory turnover. Moreover, inventory turnover is a performance metric that is evaluated *ex post* and is measured by periods (rather than real-time).

Product Variety.

We use Shannon's entropy index (Shannon and Weaver, 1949) to measure product variety in a warehouse following past studies (e.g., Straathof, 2007), which captures both the variety of SKUs and the dispersion of SKU variety. Product variety is measured as $-\sum_{i=1}^{s} p_i \ln p_i - \frac{s-1}{2n}$, where p_i is the daily proportion of SKU_i across all SKUs, s is the daily number of SKUs in the warehouse, and n is the warehouse's daily total inventory level. Prior studies have argued that the entropy index can be used to describe the extent of variety available in an environment (Starr, 1980; Gupta and Roth, 2007) and captures the uncertainty associated with products allocated in a warehouse (Smunt and Ghose, 2016). To compute the measure of product variety, we take the following steps: (1) Calculate each warehouse's daily Shannon's entropy, and (2) average the daily measures over a week to obtain the weekly warehouse's product variety.

Demand Variability.

Demand variability represents the demand-side uncertainty originating from customers and sales markets, which is beyond the individual warehouse's control. Demand variability can also be regarded as exogenous to an individual warehouse since logistics decisions cannot reversely affect the variability in customer demand. Demand variability can significantly impact inventory turnover (Gaur et al., 2005). To exclude transshipments out of demand variability, we focus on the demand related only to end-consumers and sales markets – demand from other warehouses is not considered.

Demand variability is calculated as the coefficient of variation of a warehouse's daily demand across SKUs, within a week.

Other Control Variables.

We include two control variables that may affect warehouse inventory turnover. One is a warehouse's *weekly demand* that could affect its inventory turnover. Weekly demand is calculated as the sum of a warehouse's daily demand within a week. The other control variable is the sales demand of a warehouse's *coverage* (i.e., cities), which can also impact a warehouse's inventory efficiency. We create warehouse and time (week) dummies to control for unobserved individual and time heterogeneity, corresponding to the fixed-effects model described in Section 2.5.1.

Data Transformation for Confidentiality.

Per Company A's request, we report summary statistics (Table 2.1) and correlation matrix (Table 2.2) on the standardized data of all variables, to protect sensitive information. For empirical analysis, we first logarithmize (or take a square root if the variable contains 0) and then standardize each independent variable. Using standardized independent variables in the model will not affect the consistency and validity of results. Using standardized variables can mitigate multicollinearity due to the interaction terms. For the dependent variable, we logarithmize the original values to reduce data skewness.

2.5 Model, Analysis and Results

2.5.1 Two-Stage Least-Square Fixed-Effects Model

We adopt two-stage least-square two-ways fixed-effects model (2SLS-FE) in the analysis. We choose fixed effects over random effects to account for unobserved individual warehouse heterogeneity and potential time effects. Empirically, the Hausman test supports our choice of the two-ways fixed-effects model (Chisq = 212.32^{***}). Further, the widely adopted 2SLS method is integrated with the fixed-effects model to account for the network endogeneity in the panel data. We show the first and second-stage equations of main effects for an illustrative purpose, whereas in the analysis we estimate coefficients simultaneously using `R` package `plm` (Croissant and Millo, 2018), as running two separate regressions for two stages leads to biased standard errors for β in the second-stage regression (Stock and Watson, 2011).

Equation (2.1) is the first-stage equation for direct ties. A similar equation can be constructed for transshipment and indirect ties.

$$\begin{aligned} Dir\_Tie_{it} = {} & \gamma_i^{DT} + \phi_t^{DT} + \delta_1^{DT} Lag\_Dir\_Tie_{it} + \delta_2^{DT} Lag\_Indir\_Tie_{it} \\ & + \delta_3^{DT} Lag\_Trans_{it} + \theta_1^{DT} Prod\_Var_{it} + \theta_2^{DT} Dmd\_Var_{it} \\ & + \theta_3^{DT} Wh\_Dmd_{it} + \theta_4^{DT} Cover\_Dmd_{it} + \eta_{it}^{DT} \end{aligned} \tag{2.1}$$

where the superscript DT stands for direct ties; Lag_Dir_Tie_{it}, $\text{Lag\_Indir\_Tie}_{it}$, and Lag_Trans_{it} denote the lagged direct and indirect ties and the transshipment for warehouse i at time t, respectively; γ, ϕ, δ, θ are the parameters for the model; and η_{it}^{DT} is the residual term.

The second-stage equation is:

$$\begin{aligned} Inv\_Turn_{it} = \alpha_i + \tau_t + \beta_1 Dir\_Tie_{it} + \beta_2 Inir\_Tie_{it} + \beta_3 Trans_{it} + \xi_1 Prod\_Var_{it} \\ + \xi_2 Dmd\_Var_{it} + \xi_3 Wh\_Dmd_{it} + \xi_4 Cover\_Dmd_{it} \\ + \pi^{DT}\hat{\eta}^{DT}_{it} + \pi^{IT}\hat{\eta}^{IT}_{it} + \pi^{TR}\hat{\eta}^{TR}_{it} + \varepsilon_{it} \end{aligned} \tag{2.2}$$

Where α_i and τ_t represents the unobserved individual and time heterogeneity; Dir_Tie$_{it}$, Indir_Tie$_{it}$, and Trans$_{it}$ represent direct and indirect ties and transshipment, respectively; $\hat{\eta}^{DT}_{it}$, $\hat{\eta}^{IT}_{it}$, and $\hat{\eta}^{TR}_{it}$ are the estimated residuals from the first-stage equations; and ε_{it} is the random error.

The 2SLS-FE model considering the interactions between direct ties and two variabilities can be established accordingly – we use the interactions between lagged variables and product or demand variability as instruments for the corresponding terms. The overall analysis is performed in a hierarchical manner. Specifically, we first enter the control variables (Models (1) in Table 2.3). The next model includes the main effects of direct and indirect ties to show the effect of structural embeddedness (Models (2) in Table 2.3). Finally, the analysis includes the interaction terms (Models (3) and (4) in Table 2.3).

2.5.2 Empirical Tests on Instruments

Section 2.4.2 rationalizes the instruments from a theoretical perspective. We empirically test on the instruments here. The (Wu-)Hausman test for the panel model prefers the fixed effects model, as mentioned before. Due to that we have an equal number between instruments and endogenous regressors, the Sargan test of overidentification is not necessary.

We present the weak instrument test. Multiple instruments may render the traditional weak instrument test biased (Sanderson and Windmeijer, 2016). We follow previous research and adopt the Cragg-Donald F-statistic (Cragg and Donald, 1993) that employs canonical correlations in the process. The Cragg-Donald F-statistic is 68.01^{***}, indicating all instruments are valid and strong.

2.5.3 Multicollinearity, Heteroskedasticity, Auto-correlation

We have performed additional tests on common model assumptions. We standardize the independent variables for the sake of including the interaction terms and perform the variance inflation factors (VIF) analysis. The VIFs for all variables in the analysis (including the interaction terms) are below 4, which is lower than the critical value of 10 (Kutner et al., 2005), indicating that multicollinearity is not a concern.

The studentized Breusch-Pagan test indicates (Breusch and Pagan, 1980) indicates heteroskedasticity (BP = 504.32^{***}, df = 10), while the Breusch-Godfrey/Wooldridge test for serial correlation (Breusch, 1978; Godfrey, 1978) indicates serial correlation (Chisq = 977.35^{***}). The existence of heteroskedasticity and autocorrelation may cause inconsistent standard errors of the coefficients, reducing the explanatory power (Arellano, 2003). Hence, we use the robust covariance matrix to obtain consistent standard errors.

Robust covariance matrix estimators in the style of Arellano (1987) is used because it allows a fully general structure with respect to heteroskedasticity and serial correlation. The Arellano matrix is based on nonparametric heteroskedasticity autocorrelation (HAC) covariance matrix estimators, essentially the averages of HAC estimates across individuals in the cross-section (Vogelsang, 2012). We obtain the coefficient estimates and perform Wald tests of the estimated coefficients against the heteroskedasticity-consistent Arellano covariance matrix. Corrected standard errors are robust to arbitrary autocorrelation and heteroskedasticity.

2.5.4 Results

Table 2.1 displays the summary statistics of the study variables. Table 2.2 shows the corresponding correlation matrix. From Table 2.2, inventory turnover is strongly correlated with both direct ties and indirect ties, indicating possible structural effects. The inventory turnover is also correlated with exogenous controls, justifying the inclusion of the control variables.

Table 2.1: Summary Statistics of the Variables

	Variables (N=6046)	Median	Trimmed	MAD	Min	Max	Range
1	Inventory Turnover	-0.19	-0.08	0.78	-8.64	4.45	13.09
2	Direct ties	-0.18	-0.11	1.20	-0.99	2.80	3.79
3	Indirect ties	-0.01	-0.06	1.45	-0.99	2.41	3.40
4	Transshipment	0.18	0.01	1.13	-1.88	2.42	4.30
5	Product Variety	0.06	0.13	1.14	-6.09	1.52	7.61
6	Demand Variability	-0.27	-0.16	0.67	-1.31	7.07	8.38
7	Warehouse Demand	-0.20	-0.16	0.08	-0.26	24.98	25.24
8	Coverage Demand	-0.29	-0.20	0.32	-0.64	11.32	11.95

Note: All variables are standardized (mean = 0 and sd = 1) for confidentiality.

Table 2.2: Correlation Matrix of the Continuous Variables

	Inv_turn	Dir_tie	Indir_tie	Trans	Prod_var	Dmd_var	Wh_dmd
Inv_turn							
Dir_tie	-0.35***						
Indir_tie	-0.44***	0.75***					
Trans	-0.02	0.50***	0.29***				
Prod_var	-0.49***	0.62***	0.61***	0.64***			
Dmd_var	0.15***	-0.19***	-0.17***	-0.23***	-0.29***		
Wh_dmd	-0.01	-0.02	-0.08***	0.20***	0.19***	-0.07***	
Cover_dmd	0.12***	0.11***	-0.21***	0.18***	0.04***	-0.11***	0.50***

Note: Pearson's paired test. $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ (two-tailed).

Table 2.3 shows the results of the 2SLS-FE estimation with the Arellano-robust inference. Notice that the residual sum of squares (RSS) is no longer constrained to be smaller than the total sum of squares (SST) for the 2SLS estimation (Sribney et al., 2018). Hence, R^2 and F statistic are pseudo and may not be interpretable.

We first examine the association between control variables and warehouse inventory efficiency (Models (1) in Table 2.3). Both product variety and demand variability significantly reduce inventory turnover, indicating their effects on reducing warehouse inventory efficiency. Transshipment significantly increases inventory turnover, as expected. Other control variables are also significant, justifying the inclusion of control variables. Specifically, warehouse demand is positively associated with inventory turnover, which comforts to the notion that higher demand usually leads to quicker turnover (Gaur et al., 2005).

Next, we examine the effects of direct and indirect ties. Model 2 shows a negative effect of direct ties (coeff. $= -0.244^{***}$) and a positive effect of indirect ties on inventory turnover (coeff. $= 0.143^{*}$), controlling for transshipment. The results show that increasing direct ties reduces warehouse inventory efficiency, supporting Hypothesis 1. Controlling the effect of direct ties, the increase in indirect ties is associated with higher warehouse inventory efficiency, which supports Hypothesis 2.

Models (3) and (4) in Table 2.3 show the interactions between direct ties and product and demand variabilities. We separate interaction terms into different models to further minimize multicollinearity. The interaction terms are significant. Specifically, high direct ties weaken the effect of product variety (coeff. $= 0.111^{*}$ in Model (3)), which supports Hypothesis 3. High direct ties strengthen the effect of demand variability (coeff. $= -0.024^{**}$ in Model (4)), which supports Hypothesis 4. To better understand the interaction effects, Figure 2.2 displays the interaction plots between direct ties, product variety, and demand variability. The choice of product variety and demand variability values correspond to their 25%, 50%, and 75% quantiles. We discuss implications from the interaction effects in more details in Section 2.6.

Table 2.3: Fixed-Effects 2SLS Estimation with Arellano-Robust Inference

	Dependent Variable: Inventory Turnover			
Model	(1)	(2)	(3)	(4)
Prod_var	-0.320^{***}	-0.355^{***}	-0.261^{***}	-0.353^{***}
	(0.020)	(0.028)	(0.056)	(0.028)
Dmd_var	-0.024^{*}	-0.016	-0.016	-0.029^{*}
	(0.009)	(0.010)	(0.010)	(0.011)
Wh_dmd	0.104^{***}	0.103^{***}	0.106^{***}	0.103^{***}
	(0.017)	(0.017)	(0.017)	(0.017)
Cover_dmd	-0.193^{***}	-0.201^{***}	-0.214^{***}	-0.201^{***}
	(0.016)	(0.016)	(0.018)	(0.016)
Trans	0.333^{***}	0.400^{***}	0.402^{***}	0.404^{***}
	(0.027)	(0.039)	(0.039)	(0.039)
Dir_tie		-0.244^{***}	-0.301^{***}	-0.253^{***}
		(0.053)	(0.064)	(0.053)
Indir_tie		0.143^{*}	0.154^{*}	0.146^{*}
		(0.066)	(0.066)	(0.066)
Dir_tie × Prod_var			0.111^{*}	
			(0.056)	
Dir_tie × Dmd_var				-0.024^{**}
				(0.009)
N	6,046	6,046	6,046	6,046
Pseudo R^2	0.069	0.061	0.061	0.061
Pseudo F	54.717^{***}	10.708^{***}	8.131^{***}	7.346^{***}
df for F	5; 5783	7; 5781	8; 5780	8; 5780

Note: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$
Standard errors of coefficients are displayed in the parentheses.
All independent variables are standardized due to the inclusion of the interaction.

2.5.5 Robustness Check

Results of the main analysis support all four hypotheses. We perform additional analyses as the robustness check to ensure our results are not driven by measurement or data sampling.

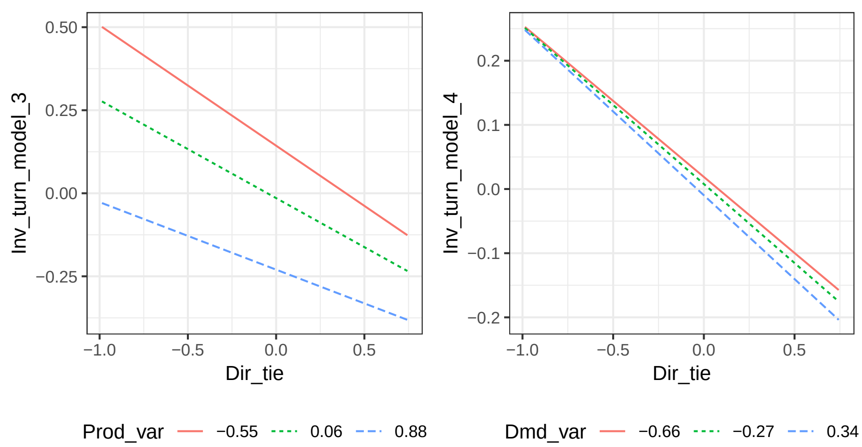


Figure 2.2: Interaction plots between direct ties and environmental uncertainty

Robustness Check on Weekly Networks.

In the main analysis, we construct the structural embeddedness measures as the averages of the daily direct and indirect ties in a week. An alternative is to construct the measures using the total count of ties within a week rather than averaging daily ties within a week. Models R1 through R3 in Table 2.4 display the results from the same 2SLS-FE model in the main analysis. Specifically, "Dir_tie_wk" and "Indir_tie_wk" denote the weekly measures. We use the corresponding lagged variables as instruments. Results indicate that both the main and the interaction effects are significant and consistent with the main analysis. The similar magnitudes of the coefficients ensure that the results do not vary with alternative measurements.

Robustness Check on Temporary Warehouses.

Through descriptive analysis and visualization of the data, we notice a few temporary warehouses in the network – they exist for a couple of months (usually within 3 months). To ensure that our results are not driven by the temporary warehouses, we create a reduced dataset without those warehouses. The reduced data have $5,876$

Table 2.4: Robustness Check: Variable Measure (R1-R3) and Data Sampling (R4-R6)

	Dependent Variable: Inventory Turnover					
Model	(R1)	(R2)	(R3)	(R4)	(R5)	(R6)
Prod_var	-0.351^{***}	-0.173^{***}	-0.348^{***}	-0.345^{***}	-0.273^{***}	-0.342^{***}
	(0.028)	(0.052)	(0.028)	(0.025)	(0.046)	(0.025)
Dmd_var	$-0.019^{\dagger}$	$-0.017^{\dagger}$	-0.029^{**}	-0.002	-0.002	-0.013
	(0.010)	(0.010)	(0.011)	(0.010)	(0.010)	(0.011)
Wh_dmd	0.104^{***}	0.108^{***}	0.103^{***}	0.101^{***}	0.102^{***}	0.100^{***}
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Cover_dmd	-0.200^{***}	-0.220^{***}	-0.202^{***}	-0.202^{***}	-0.212^{***}	-0.202^{***}
	(0.017)	(0.018)	(0.017)	(0.016)	(0.017)	(0.016)
Trans	0.382^{***}	0.389^{***}	0.387^{***}	0.405^{***}	0.407^{***}	0.409^{***}
	(0.040)	(0.041)	(0.040)	(0.034)	(0.035)	(0.034)
Dir_tie_wk	-0.149^{*}	-0.228^{**}	-0.159^{*}			
	(0.061)	(0.071)	(0.062)			
Indir_tie_wk	0.141^{*}	0.158^{*}	0.144^{*}			
	(0.070)	(0.070)	(0.070)			
Dir_tie_wk × Prod_var		0.179^{***}				
		(0.046)				
Dir_tie_wk × Dmd_var			-0.020^{**}			
			(0.008)			
Dir_tie				-0.285^{***}	-0.324^{***}	-0.290^{***}
				(0.050)	(0.056)	(0.050)
Indir_tie				0.145^{*}	0.154^{*}	0.147^{*}
				(0.062)	(0.063)	(0.062)
Dir_tie × Prod_var					$0.084^{\dagger}$	
					(0.046)	
Dir_tie × Dmd_var						-0.024^{**}
						(0.008)
N	6,046	6,046	6,046	5,876	5,876	5,876
Pseudo R^2	0.056	0.059	0.056	0.061	0.061	0.061

Note: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$
Standard errors of coefficients are displayed in the parentheses.
All independent variables are standardized due to the inclusion of the interaction.

entries with 160 warehouses along 80 weeks.

We apply the same 2SLS-FE model and report results as Models R4 through R6 in Table 2.4. To differentiate from the first robustness check, we show related main and interaction effects in separate rows. All the effects of interest are significant except for the marginally significant interaction between direct ties and product variety. The directions and magnitudes of the effects are consistent with the main results. Hence, our results are robust to the issue of data sampling due to the temporary warehouses.

2.6 Discussion

2.6.1 Theoretical Implications

Essay 1 contributes to the growing empirical literature that investigates the impacting factors of inventory performance across different settings using longitudinal data (e.g., Gaur et al., 2005; Gaur and Kesavan, 2008; Kolias et al., 2011; Rajagopalan, 2013; Lee et al., 2015). Using a proprietary panel dataset, we extend this empirical literature stream by showing the influences of warehouse structural embeddedness, product variety, and demand variability on warehouse inventory efficiency, while past empirical studies often investigate the relationships between firm-level factors (e.g., gross margin, capital intensity, firm size) on inventory performance in a specific industry (e.g., retailing).

The main effects suggest that a warehouse's ego-network structure exerts differing effects on warehouse inventory efficiency. Specifically, the results show that a focal warehouse with few direct connections and high interconnectedness (the neighboring warehouses have many connections among themselves) with respect to its ego network can achieve high warehouse inventory efficiency. We transform the coefficients back to their original scales and find that, holding the control factors constant, one more direct tie is associated with 6.78% reduction in inventory turnover and a 1% increase in indirect ties (proportional density) results in a 1.61% increase in inventory turnover. In sum, our results demonstrate that a focal warehouse with few direct ties and high interconnectedness in its ego-network exhibit better inventory efficiency.

Additionally, at the warehouse level, we show that demand variability and product variety interact with network structural factor and jointly affect inventory efficiency significantly. As predicted by past studies, demand variability influences warehouse inventory efficiency negatively. Further, we find that the extent of direct ties strengthens the negative effect of demand variability, which suggests that a busy warehouse with many direct ties and faced with high demand variability cannot expect to perform at the same level as its less-connected counterparts in terms of inventory efficiency. From an individual warehouse's perspective, a logistics strategy to potentially mitigate high demand variability is to have relatively fewer direct connections to maintain stability of internal warehouse operations. In a sense, at the warehouse level, reducing demand variability through other means that have discussed in the literature (e.g., better forecast through collaboration and information sharing) could be a better approach to increase inventory efficiency.

The findings of product variety also provide an interesting perspective regarding warehouse inventory management. Our results show that the extent of direct ties can mitigate the negative effects of produce variety to a certain extent (see Figure 2.2). Based on our findings, firms should allocate high extent of product variety (various and disperse SKUs) to the warehouses with more direct ties. Though increasing direct ties can increase coordination tasks for a warehouse, we find that having more direct ties is an effective way to handle increasing varieties of products. In a sense, our results help explain the e-commerce giants' expanding decision regarding their warehouse network observed in recent headlines. Diversity in end consumers' tastes are demanding high product variety (Lancaster, 1990) and the e-commerce companies are responding to such demand by offering increasing numbers of SKUs as a strategic move to boost sales. Nonetheless, product variety has been shown to increase inventory level and reduces inventory turnover in the literature (Zipkin, 2000). From an e-commerce firm's standpoint, building a more complex warehouse networks (i.e., having more direct ties and indirect ties within warehouse network) not only helps increase customer responsiveness but also helps alleviate the potential drawback of high product variety at the individual warehouse level.

2.6.2 Managerial Implications

This study provides two aspects of managerial implications with respect to the assessment and improvement of warehouse inventory management. On the one hand, assessing warehouse inventory efficiency without accounting for the related structural factors can lead to misguided results. Based on our analysis, a warehouse could experience high or low inventory efficiency partly due to its structural embeddedness in the network. Therefore, when conducting warehouse network planning or performance evaluation, managers should consider a warehouse's structural position in the network. Naïvely benchmarking warehouse's inventory efficiency without considering the effects of structural embeddedness will likely lead to biases, misleading assessments, and wrong expectations.

On the other hand, this study suggests ways to mitigate environmental uncertainties through network structural decisions. Demand variability can make inventory management more difficult for structurally more central warehouses. As a result, managers should be cautious about inventory turnover of the warehouse embedded centrally in a hub-spoke network (i.e., high direct ties, low indirect ties) that experiences high demand variability, since such structure is not well suited for high demand variability and could greatly reduce warehouse inventory efficiency. On the contrary, the results indicate that a warehouse with the preferred network structure (low direct ties, high indirect ties) can better handle high demand variability – managers could consider assigning popular products with high demand variability to structurally more peripheral warehouses that have the preferred network structure.

In contrast, retailers should consider allocating or storing more product varieties or higher dispersion of SKUs into a warehouse that is centrally embedded in the hub-spoke network (i.e., high direct ties, low indirect ties). For structurally more peripheral warehouses, storing or allocating low product variety (maybe a few kinds of popular products) can be better for warehouse inventory efficiency.

2.7 Conclusion and Future Research

In Essay 1, we examine the effects of direct and indirect ties – two aspects of structural embeddedness – in the setting of a warehouse network. We find that direct ties negatively affect while indirect ties positively affect the efficiency of warehouse inventory management. We also find that direct ties interact with product variety and demand variability, which provides a better understanding to warehouse managers with regards to performance implications of individual warehouses and how network structure can strength or weaken the effects of product variety and demand variability.

This study is not without limitations, which provide future research opportunities. First, this study examines the effects of structural embeddedness on the *individual* warehouse's inventory performance. In other words, the analysis and implications are mostly applicable to individual warehouses instead of the entire warehouse network. Future research may focus on the entire network and improve the system-wide (network-wide) efficiency performance. Future research could focus on examining preferred structural properties from the whole network's perspective and provide a more fine-grained understanding regarding network configuration and the entire network's inventory management.

Second, we examine one warehouse network of one specific product type. Although the results are robust to the variable measure and data sampling, future research can include networks in different contexts to re-examine the structural effects, to either confirm or reject out findings. In addition, researchers may also explore how the warehouse network evolves (changes the structure) overtime, and how each warehouse makes logistics decisions to adapt to some performance criteria.

Finally, we have focused on the efficiency of warehouse inventory management in this study, as most online retailers maintain certain customer service levels for warehouses. Future research could discuss the trade-off between efficiency and effectiveness of warehouse inventory management as two strategic focuses (Heikkilä, 2002; Lee, 2004). In the context of warehouse network, warehouse efficiency and effectiveness are two performance goals (Ecklund, 2010). We argue in this study that, for one retailer company, the fill rate that captures effectiveness may usually be the same across the warehouse network (which is the case based on our field study). While we focus on warehouse inventory efficiency that stands out as a differentiating performance metric, the aspect of effectiveness should be concerned as well. In fact, a warehouse might be efficient, but if the service level is not achieved, it is ineffective. Hence, future research may investigate into the effect of structural embeddedness on warehouse effectiveness (stockout rate can be a proxy as indicated by Beamon (1998)). Efficiency and effectiveness are mutually affected. Hence, sophisticated specification techniques may be required to estimate simultaneous equations.

At closing, we hope that this study generates more interests in examining warehouse inventory performance from a network structural perspective, which has been largely neglected in the current literature.

Chapter 3

Where to Improve Resilience in the Supply Chain Network under Stochastic Disruptions?

3.1 Introduction

Over the past few decades, firms have recognized the importance of managing supply chain disruptions (Hendricks and Singhal, 2005). Scholars have developed various frameworks to guide firms on *what* capabilities they need to develop to improve resilience (e.g., Pettit et al., 2013). However, *where* to make these investments in the firm's supply chain network has received less attention. "While many consultants, researchers, and managers agree on the importance of supply chain resilience, there is less agreement on ... where to invest to mitigate risk and recover from disruptions – to shape and influence resiliency" (Melnyk et al., 2015). In the past, supply chain network complexity and data limitations created obstacles to assess where firms can best invest in their supply chain network to improve resilience. But, increasing levels of rich data across the supply chain network provide an opportunity to develop data-driven investment strategies to improve resilience in the network. This research takes a data-driven prescriptive analytics approach to investigate the following research question: *Where to invest limited resources in the supply chain network to improve supply chain resilience?*

We investigate the optimal investment strategy in resilience by considering a general directed acyclic supply chain network (e.g. a distribution network) – a centralized system where the nodes (facilities) in the network experience stochastic independent disruptions (e.g. fire, strike, security issue). Each node has a limited capacity to process material flows (e.g. to distribute), and a disruption reduces the node's capacity. A decision-maker needs to determine which nodes to invest in to make the network most resilient. This investment reduces the node's probability of a disruption. The decision-maker has limited resources and makes binary decisions about which node to invest in to improve network resilience. Making investments in a node's resilience should result in higher expected total material flows through the supply chain network. The investment strategy depends in part on the routing mechanism of material flow through the network. A network can have a *nonreroutable* mechanism where material can flow only through pre-specified paths, or it can have a *reroutable* mechanism where material can flow through alternative paths (Carey and Hendrickson, 1984). The mechanism used in a supply chain network depends on factors such as contracts and the nature of the materials (products).

Scholars have debated what nodes are critical to the overall supply chain network resilience. Some have argued that the decision-makers should identify the *critical nodes* to improve resilience (Craighead et al., 2007), while others argue that they should identify the nodes that reside on a *critical path* (Nair and Vidal, 2011) to improve resilience. That is, some take a node perspective to understand resilience in a supply chain network, while others have argued for a path perspective. Although these scholars did not explicitly investigate investment in resilience, this debate has potential implications for the investment decision.

This research investigates the investment decision in supply chain network resilience as a two-stage stochastic optimization problem. The analysis shows that the probability of a disruption and the routing mechanism influence whether the decision-maker should take a node or a path investment strategy. The results indicated that a node investment strategy works best for a supply chain network where the nodes face rare disruptions, while a path investment strategy works best when nodes face frequent disruptions. In addition, the investment in a supply chain network with a nonreroutable flow mechanism is *monotone supermodular* – hence, a greedy algorithm works well. But, the investment in a supply chain network with the reroutable flow mechanism or where the nodes face a moderate level of disruptions is complicated and intractable. This suggests that determining the optimal investment strategy depends on the characteristics of the supply chain network.

Data-driven prescriptive analytics plays a vital role in determining the investment strategy for a supply chain network with a reroutable flow mechanism and/or a moderate level of disruptions. To further investigate the general investment strategies, we collaborate with a leading supply chain management company (hereafter "Company A") that faces the problem of where to invest their limited resources to improve their network resilience. From Company A's supply chain network data, we find that node capacity, average path length, and the node flow centrality are critical factors that interact with the probability of disruptions and the routing mechanism to affect the overall network resilience. In general, the findings show that when applied to more realistic networks, the optimal investment should never follow a pure node or path investment strategy. As a result, data-driven prescriptive analytics is critical to determine the best investment strategy.

This study makes three contributions to the literature of supply chain disruptions. First, to the best of our knowledge, it is the first data-driven prescriptive work that directly characterizes optimal strategies of resilience investment. Second, this study highlights the network perspective and shows that network structural factors affect investment strategies on network resilience. This echoes Kim et al. (2015, p.56)'s argument that "node criticality needs to be understood in terms of the overall network structure". We show that a node's capacity plays a critical role in a scenario of rare disruptions, but its structural position in the network gains a higher level of prominence in other scenarios. Third, we clarify the contingency about disruption probability and routing mechanism when making investments. Our findings correspond well with the notion that "the nature of the disruptions is a key determinant of the optimal strategy" (Tomlin, 2006, p.639).

This study also contributes to practice. Our findings and the greedy algorithm can significantly facilitate decision making to improve the resilience of a supply chain network. The problem is nontrivial because the search of the optimal solution grows exponentially with the size of the network (especially for realistic networks). With rich data, the proposed prescriptive analytics guides managers to assess conditions in advance, sense the weights of nodes or paths (e.g., product lines), and allocate resilience resources accordingly. Moreover, managers need to take a system view when thinking about improving resilience, rather than focusing on isolated nodes. To achieve the highest effectiveness of resilience investment, managers should incorporate the path perspective into the node-level investment. Finally, the model and findings in this study are applicable to various networked systems, including supply networks of finished products, assembly networks or production systems, and distribution networks. While the flow network naturally describes the operations of distribution networks and networks of finished products, the normalized flow amounts in the model can imply the success probability of an assembling or production, although the addition of the original flow amounts are practically meaningless in those systems.

The rest of Essay 2 has the following organization. Section 3.2 gives a literature review on supply chain resilience from a network perspective. Section 3.3 gives the model formulation, while Section 3.4 analytically characterizes the optimal strategies of resilience investment under extreme probabilities of disruptions. Section 3.5 adopts the data-driven analytics by using Company A's network and operational data and prescribes optimal investment strategies in general situations. Section 3.6 discusses the findings, managerial implications, and future research.

3.2 Literature Review on Supply Chain Network Resilience

A supply chain network consists of nodes that process and store material, and arcs that transport material between the nodes (Borgatti and Li, 2009; Carter et al., 2015; Lee and Billington, 1993). Scholars have proposed various frameworks and strategies to improve supply chain resilience (e.g., Chopra and Sodhi, 2014; Christopher and Peck, 2004; Kleindorfer and Saad, 2005; Sheffi and Rice Jr, 2005; Simchi-Levi et al., 2014, 2015; Tang, 2006). They have also noted that firms should take a network perspective when managing supply chain disruptions, because resilience depends, in part, on the network structure (Kim et al., 2015). Research shows that the design and structure of a supply chain network influences resilience (Snyder and Daskin, 2007).

Much of the emerging literature on supply chain network resilience has focused on *how* to reduce disruption and improve resilience. For instance, Pettit et al. (2013) identified several capabilities that firms need to develop to improve resilience. They argued that a firm's capabilities should align with their disruption vulnerabilities. Simchi-Levi et al. (2014) and Simchi-Levi et al. (2015) proposed how to identify risks and mitigate different kinds of disruptions. These studies offer valuable insights and provide strategies on *how* to improve resilience.

Emerging studies have begun to take a network structure perspective to examine *how* firms can improve resilience. This perspective takes into consideration the firm's position in the network and/or the supply chain network topology. These studies showed that the average path length (Nair and Vidal, 2011) and other network properties (e.g., clustering coefficient, scale-free and small-world characteristics) (Albert et al., 2000; Basole and Bellamy, 2014; Kim et al., 2015) can significantly influence risk mitigation. Taking a network perspective, firms can make optimal sourcing decisions based on their positions in the network under disruption risks (Ang et al., 2016; Bimpikis et al., 2017, 2018). Firms can also make optimal investment to mitigate disruptions, as a response to their suppliers' investments (Bakshi and Mohan, 2015). Despite the growing literature on *how* to improve resilience, research on *where* to improve resilience in a supply chain network is limited. However, determining *where* to make improvements is critical within the context of a supply chain network.

Some studies outside of the supply chain management area have investigated *where* to protect in the network. For instance, the network interdiction problem has its origins in the military (Golden, 1978), and examines how to maximize the flow through a network in the face of disruptions (e.g., Cormican et al., 1998). This may be due to terrorism which is critical to maintain national infrastructure such as electric power systems (Salmeron et al., 2004). Similarly, some have focused on investments in the national highway system to mitigate intentional disruption threats (e.g., Peeta et al., 2010), which is critical to managing the national infrastructure. However, our study investigates random disruptions, often due to natural disasters (unintentional) rather than intentional threats within the context of a supply chain network.

A disruption in the supply chain network effectively reduces the supplier's capacity. Several studies (e.g., Wang et al., 2010, 2014) have developed models for capacity and yield uncertainties to improve supplier performance, but have not considered the broader network perspective. The works of Wallace (1987) and Wollmer (1991) most closely connects to our research. They investigated *where* (mostly which edges) to protect in the network to maximize the overall flow under uncertainty. Despite the similarity of their work, our model setup and prescribed strategies are fundamentally different from theirs. We make the following contributions. First, we consider investment in nodes to improve resilience. In practice, many companies, such as Honda, Toyota, BMW, Intel, and Siemens, devote significant resources to improving the reliability of their suppliers' facilities and restoring their capacities should a disruption occur (Handfield et al., 2010; Liker and Choi, 2004). Second, we consider both nonreroutable and reroutable supply chain networks. Our reroutable model is a two-stage stochastic program with decision-dependent uncertainty (DDU, or endogenous uncertainty, in which uncertainty depends on the decision variables), where the first-stage decisions affect the actual probabilities. Limited research has studied this kind of model (Medal et al., 2016). Third, we prescribe optimal strategies in a data-driven manner (optimal investment is used as the data source for the data-driven prescriptions) (Simchi-Levi, 2013), because the problem is proven to be analytically hard to solve in general situations. Most previous studies (e.g., Cormican et al., 1998; Wallace, 1987; Wollmer, 1991), although proposed efficient (near-optimal) algorithms, focused on the algorithm efficiency but did not characterize the solutions.

In general, Essay 2 studies a centralized system (i.e., the supply chain network) in which a focal firm needs to best fortify existing nodes against risks to achieve the maximized yield. This research corresponds well with the acknowledgment that improving supply chain resilience should look beyond the focal company's facilities and understand the entire supply chain network (Kim et al., 2015).

3.3 Model Formulation

This section formulates the analytical model used to identify strategies for resilience investment. We describe the supply chain network and the *routing mechanism* of material flows under disruptions. A two-stage stochastic combinatorial optimization model is developed to determine the optimal investment in resilience.

3.3.1 Supply Chain Network

The supply chain network under examination can be a focal firm's supply base or a distribution/logistics network, which can be mapped out of the rich data. In the model formulation, we investigate a generic supply chain network that is a directed acyclic graph with physical material flows, denoted as $G = (V, E)$ with a source $v_z \in V$, a sink $v_0 \in V$, a mapping $\omega\colon V \to \mathbb{R}_{\geq 0}$ that represents capacities of the nodes, and a mapping $w\colon E \to \mathbb{R}_{\geq 0}$ that represents capacities of the edges. The source and/or sink may be dummy if multiple actual sources and/or sinks exist. A flow network with the directed acyclic structure occurs commonly in practice such as the automotive supply chains and the distribution systems, and is also well established in the literature (e.g., Magnanti et al., 2006). Specifically, V is the set of nodes (firms or facilities) and $E \subseteq V \times V$ is the set of directed edges that are ordered pairs of nodes. Let $V_S = V \backslash \{v_z, v_0\}$ be the set of supply chain nodes $1, 2, \ldots, I$ ($|V_S| = I$) that are vulnerable to random disruptions and can be invested on.

Materials flow on the paths through the supply chain network. A *path* is a sequence of connected nodes and edges that originate at v_z and terminate at v_0. Let $\phi_k \in \mathbb{R}_{\geq 0}$ be the amount of material flow along the path $k \in \Phi$ (the "path flow"), where Φ is the set of all paths in the graph G. The path flows through nodes and edges are subject to their capacities.

$$\sum_{k \in \Phi | i \in k} \phi_k \leq \omega_i, \quad \forall i \in V, \quad \text{and/or} \tag{3.1}$$

$$\sum_{k \in \Phi | (i,j) \in k} \phi_k \leq w_{ij}, \quad \forall (i,j) \in E \tag{3.2}$$

where $i \in k$ and $(i, j) \in k$ mean that node i and edge (i, j) are on path k, respectively.

Contextually, node capacity ω_i enables node $i \in V_S$ to process received materials from upstream and distribute them out to downstream. Edge capacity w_{ij} indicates

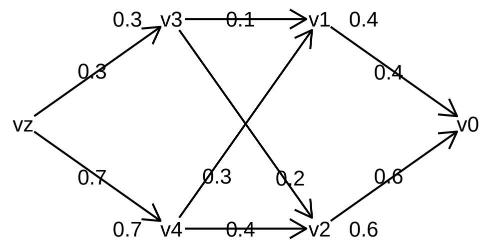


Figure 3.1: An example of the supply chain network with node and edge capacities

the potential material quantity in the paired transaction. Although node and edge capacities can be arbitrary non-negative numbers, in a centralized network where the manager can assess the product lines or the distribution routes, it is equivalent to examine capacities that are reduced to the following characterization without loss of generality.

$$\omega_i = \sum_{k \in V} w_{ki} = \sum_{j \in V} w_{ij}, \quad \forall i \in V_S \tag{3.3}$$

In the sense of Equation (3.3), every node prepares its normal capacity equal to the anticipated demand. All quantities are scaled to the same unit so that they are comparable and additive. The amount of materials sent by v_z is normalized as $\omega_z = \sum_{i \in V_S} w_{zi} = 1$. Correspondingly, $\omega_0 = \sum_{i \in V_S} w_{i0} = 1$. For the distribution network and the supply network of finished products, the total maximum flow through the network ("max-flow") represents the available proportion of the supply, while for the assembly network that has an either-or outcome, the normalized max-flow may represent the success or completion probability of assembling or production.

Figure 3.1 illustrates a supply chain network with node capacities (beside the nodes) and edge capacities (on the edges). This example will be used to illustrate the supply chain network operations and resilience investment throughout the rest of this section.

3.3.2 Disruption and Resilience Investment

Disruptions can occur stochastically yet independently on node $i \in V_S$ with a probability of $p_i^D \in [0, 1]$. p_i^D equivalently represents the fraction of time that node i is disrupted, hence also known as the "disruption frequency". The realized disruption reduces the node capacity. Hence, we use a binary random variable $L_i, i \in V_S$ as the capacity scalar with respect to the disruption on i, where L_i takes values l and 1 with probabilities p_i^D and $1 - p_i^D$, respectively. $l \in [0, 1)$ indicates the remaining proportion of the capacity after disruption and is presumably identical across V_S. To avoid trivial cases, $L_z = L_0 \equiv 1$, i.e., the source and the sink are never disrupted.

The disruption probability p_i^D reflects the strategies of disruption mitigation and resilience improvement such as capacity buffer and security training. The resilience investment on node $i \in V_S$ (e.g., more buffer and training) can reduce i's probability of disruption by a pre-specified $\delta_i \in (0, 1]$ (known as the "investment benefit"). Let $\mathcal{N} \subseteq V_S$ be the set of invested nodes. Let x_i be a binary indicator of the investment decision on $i \in V_S$ (1 if $i \in \mathcal{N}$ and 0 otherwise). Therefore, the realization of the capacity scalar L_i depends on the following cases:

$$L_i = \begin{cases} 1 & \text{with } 1 - p_i^D + p_i^D \delta_i x_i \\ l & \text{with } p_i^D - p_i^D \delta_i x_i \end{cases} \tag{3.4}$$

Under limited resources of resilience investment, the supply chain manager can invest on at most K nodes, with presumably the same investment cost across nodes. Hence, $|\mathcal{N}| \leq K$. Managers want to invest on the nodes at the beginning to most effectively improve the resilience of the supply chain network against subsequent disruptions (in finite or infinite time steps). In this sense, the investment decision occurs at the first stage, in order to maximize the *expected* max-flow under uncertainty at the second stage. Let y be the outcome of interest (i.e., the maximized expected max-flow). We therefore obtain:

$$y = \max_{\boldsymbol{x}} \quad \mathbb{E}h(\mathcal{N}), \quad \forall \mathcal{N} \subseteq V_S \tag{3.5}$$

$$\text{s.t.} \quad \sum_{i \in V_S} x_i \leq K \tag{3.6}$$

$$x_i = \begin{cases} 1 & \text{if } i \in \mathcal{N} \\ 0 & \text{if } i \notin \mathcal{N} \end{cases} \tag{3.7}$$

where h denotes the max-flow that is a function of the investment set. h is subject to the realization of random disruptions, determined by the disruption probability and the investment benefit. For notational simplicity, we let $\boldsymbol{x} \in X$ denote the satisfaction of the first-stage constraints of limited resource and binary choice, i.e., constraints (3.6) and (3.7). Throughout the rest of Essay 2, "$\boldsymbol{x} \in X$" will be used underneath the maximization symbol without repeating the constraints.

3.3.3 Max-Flow and Routing Mechanism

Below we characterize the max-flow set function h with respect to specific realization of disruption. h can be affected by the routing mechanism of path flows. Depending on whether path flows affect each other, a flow can be *nonreroutable* or *reroutable*.

Nonreroutable Flow.

Nonreroutable flow refers to the situation where the path flows in the supply chain network cannot be adjusted *ex post* in response to a disruption to increase the total flow through the network. In a real-world situation, nonreroutable flow may be due to things like contracts, regulations (e.g., FDA), product characteristics, or shipping requirements. In this situation, any node disruption on the path reduces the path flow ϕ_k to $l \times \phi_k$. Figure 3.2 shows one realization of disruptions ($l = 0.2$) and the

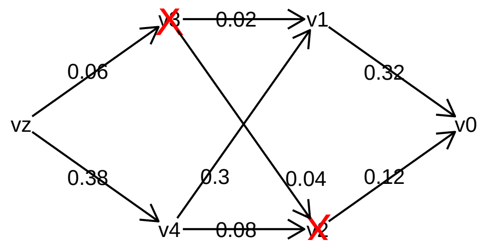


Figure 3.2: Disruptions on v2 and v3 and nonreroutable flows through edges

Table 3.1: Max-flow for nonreroutable flow under disruptions ($l = 0.2$)

Disruption	Max-flow	Disruption	Max-flow	Disruption	Max-flow
None	1	1,2	0.2	1,2,3	0.2
1	0.68	1,3	0.52	1,2,4	0.2
2	0.52	1,4	0.36	1,3,4	0.2
3	0.76	2,3	0.44	2,3,4	0.2
4	0.44	2,4	0.28	1,2,3,4	0.2
		3,4	0.2		

corresponding (max) flows through edges under nonreroutable flow. The max-flow is 0.44. Table 3.1 gives all possible max-flows under $l = 0.2$.

Under nonreroutable flow, the normal path flows may not change once determined (i.e., without recourse), regardless of disruption realization. The path flow remains its normal level only when all nodes on the path are healthy and reduces to $\phi_k \cdot l$ otherwise. Therefore, the expectation of a path flow can be determined with respect to node disruption probability and investment benefit.

$$
\begin{aligned}
\mathbb{E}\phi_k &= \phi_k \cdot \prod_{i \in k} (1 - p_i^D + p_i^D \delta_i x_i) + \phi_k \cdot l \cdot [1 - \prod_{i \in k} (1 - p_i^D + p_i^D \delta_i x_i)] \\
&= (1 - l)\phi_k \cdot \prod_{i \in k} (1 - p_i^D + p_i^D \delta_i x_i) + \phi_k l
\end{aligned}
\tag{3.8}
$$

The optimal resilience investment maximizes the expected max-flow (Equation

(3.5)), which equals the sum of expected path flows under nonreroutable mechanism.

$$\mathbb{E}h^N(\mathcal{N}) = \sum_{k\in\Phi} \mathbb{E}\phi_k \tag{3.9}$$

where the superscript N on h denotes the nonreroutable situation. Model 1 gives the problem of optimal investment for nonreroutable flow (with superscript N on y). Particularly, path flows are specified prior to disruptions and are constrained by the edge capacities.

Model 1 (Problem of optimal investment for nonreroutable flow)

$$y^N = \max_{\boldsymbol{x}\in X, \boldsymbol{\phi}\in\mathbb{R}^{|\Phi|}_{\geq 0}} \quad l + (1-l)\sum_{k\in\Phi}\phi_k \prod_{i\in k}(1 - p_i^D + p_i^D\delta_i x_i) \tag{3.10}$$

$$\text{s.t.} \quad \sum_{k\in\Phi|(i,j)\in k} \phi_k \leq w_{ij}, \quad \forall(i,j)\in E \tag{3.11}$$

Reroutable Flow.

Reroutable flow refers to the situation where the path flows in a supply chain network can be adjusted *ex post* in response to a disruption to increase the total flow through the network. In other words, the remaining node capacity after disruption can be used flexibly across downstream nodes to maximize the entire max-flow. In a real-world situation, facilities can collaborate, for instance, using nonexclusive assets. In a practical sense, the edge capacity becomes less restrictive – node i is able to send more than w_{ij} to node j given realized disruptions as long as the flow through i is bounded by the realized node capacity. Figure 3.3 shows the same realization of disruptions as Figure 3.2 and the corresponding node capacities given $l = 0.2$. The max-flow becomes 0.52. Table 3.2 similarly exhausts all max-flows under $l = 0.2$, with respect to reroutable flow.

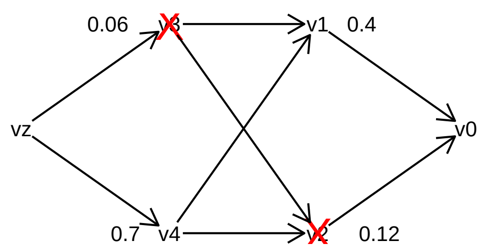


Figure 3.3: Disruptions on v2 and v3 and reroutable flows through nodes

Table 3.2: Max-flow for reroutable flow under disruptions ($l = 0.2$)

Disruption	Max-flow	Disruption	Max-flow	Disruption	Max-flow
None	1	1,2	0.2	1,2,3	0.2
1	0.68	1,3	0.68	1,2,4	0.2
2	0.52	1,4	0.44	1,3,4	0.2
3	0.76	2,3	0.52	2,3,4	0.2
4	0.44	2,4	0.44	1,2,3,4	0.2
		3,4	0.2		

Model 2 gives the maximized expected max-flow with resilience investment $\mathcal{N}$ for reroutable flow. The superscript R on y and h denotes the reroutable situation. Notice that the path flows can be adjusted *ex post* and are constrained by the node capacities. Hence, Model 2 is a two-stage stochastic combinatorial optimization with recourse (Birge and Louveaux, 2011). Table 3.3 presents the notation used.

Model 2 (Problem of optimal investment for reroutable flow)

$$y^R = \max_{\boldsymbol{x} \in X} \qquad \mathbb{E} h^R(\mathcal{N}), \quad \forall \mathcal{N} \subseteq V_S \tag{3.12}$$

where

$$h^R(\mathcal{N}) = \max_{\boldsymbol{\phi} \in \mathbb{R}_{\geq 0}^{|\Phi|}} \qquad \sum_{k \in \Phi} \phi_k \tag{3.13}$$

$$\text{s.t.} \qquad \sum_{k \in \Phi | i \in k} \phi_k \leq L_i \omega_i, \quad \forall i \in V \tag{3.14}$$

We illustrate the resilience investment models using the previous network example. Suppose an identical probability of disruption $p_i^D = 0.4$ and identical investment benefit $\delta_i = 0.5$ across nodes. The supply chain manager has resources to invest on at most $K = 2$ nodes. Table 3.4 displays $C_4^2 = 6$ resilience investment options and the corresponding expected max-flows for both routing mechanisms. Specifically, columns of h^N and h^R come from Tables 3.1 and 3.2. Columns of $\Pr(i, j)$ displays the probabilities of the corresponding disruptions given the investment on nodes i and j.

Table 3.3: Notation in the model

Sets	Description
V	Set of nodes in the graph G, $V = \{v_0, 1, \ldots, I, v_z\}$
V_S	Set of supply chain nodes, i.e., investable nodes, $V_S = V \backslash \{v_0, v_z\}$
E	Set of edges in the graph G
X	Set of feasible investment satisfying constraints (3.6) and (3.7)
Φ	Set of paths in the graph G
Parameters	**Description**
I	Number of supply chain nodes (investable nodes). $\lvert V_S \rvert = I$
K	Maximal number of nodes that can be invested
l	Remaining proportion of the capacity should disruption occur
L_i	Binary random capacity scalar to indicate disruption
p_i^D	Probability of disruption on node i without investment
w_{ij}	Normal capacity for edge (i, j) without disruption
ω_i	Normal capacity for node i without disruption
ϕ_k	Amount of material flow along the path k
δ_i	Investment benefit in reducing the probability of disruption
Decision	**Description**
$\mathcal{N}$	Set of invested nodes to be determined. $\mathcal{N} \subseteq V_S$
x_i	Investment indicator corresponding to $\mathcal{N}$. $x_i \in \{1, 0\}$

According to Table 3.4, under $(p_i^D, \delta_i, K, I) = (0.4, 0.5, 2, 4)$, the optimal resilience occurs by investing on nodes 2 and 4 for both routing mechanisms. Even in this simple example, it still takes considerable computation to identify the optimal strategy. In fact, the entire space for the different node states (i.e., being disrupted or healthy)

Table 3.4: Resilience investment under $K = 2$ and expected max-flows

Disruption	h^N	h^R	Pr(1,2)	Pr(1,3)	Pr(1,4)	Pr(2,3)	Pr(2,4)	Pr(3,4)
None	1	1	0.2304	0.2304	0.2304	0.2304	0.2304	0.2304
1	0.68	0.68	0.0576	0.0576	0.0576	0.1536	0.1536	0.1536
2	0.52	0.52	0.0576	0.1536	0.1536	0.0576	0.0576	0.1536
3	0.76	0.76	0.1536	0.0576	0.1536	0.0576	0.1536	0.0576
4	0.44	0.44	0.1536	0.1536	0.0576	0.1536	0.0576	0.0576
1,2	0.2	0.2	0.0144	0.0384	0.0384	0.0384	0.0384	0.1024
1,3	0.52	0.68	0.0384	0.0144	0.0384	0.0384	0.1024	0.0384
1,4	0.36	0.44	0.0384	0.0384	0.0144	0.1024	0.0384	0.0384
2,3	0.44	0.52	0.0384	0.0384	0.1024	0.0144	0.0384	0.0384
2,4	0.28	0.44	0.0384	0.1024	0.0384	0.0384	0.0144	0.0384
3,4	0.2	0.2	0.1024	0.0384	0.0384	0.0384	0.0384	0.0144
1,2,3	0.2	0.2	0.0096	0.0096	0.0256	0.0096	0.0256	0.0256
1,2,4	0.2	0.2	0.0096	0.0256	0.0096	0.0256	0.0096	0.0256
1,3,4	0.2	0.2	0.0256	0.0096	0.0096	0.0256	0.0256	0.0096
2,3,4	0.2	0.2	0.0256	0.0256	0.0256	0.0096	0.0096	0.0096
1,2,3,4	0.2	0.2	0.0064	0.0064	0.0064	0.0064	0.0064	0.0064
$\mathbb{E}h^N$ (nonreroutable)			0.584	0.5584	0.6032	0.5808	**0.6256**	0.584
$\mathbb{E}h^R$ (reroutable)			0.602	0.583	0.625	0.602	**0.650**	0.602

has a cardinality of 2^I and the search space has a cardinality of C_I^K. Therefore, facing a realistic network with a large I and a reasonably big K, managers urgently need optimal strategies to facilitate practical decision making of resilience investment. The following sections characterize and prescribe such optimal strategies.

3.4 Optimal Resilience Investment Strategies

This section analyzes the model and characterizes optimal strategies for resilience investment under various levels of disruption probabilities. Throughout the rest of Essay 2, we assume a common $p^D \in [0, 1]$ across nodes and a common $\delta > 0$ for investment benefit. At the beginning of this section, we highlight the analytical findings and direct readers to specific subsections.

1. Given the same network and level of investment, the reroutable mechanism yields a larger or equivalent maximized expected max-flow than nonreroutable. (Proposition 1 in Section 3.4.1)

2. Under rare disruptions, investment priority should be given to those nodes with the highest capacities regardless of the routing mechanism. (Proposition 2 and Corollary 1 in Section 3.4.2)

3. Under frequent disruptions, investment priority should be given to the entire paths, both for nonreroutable flow and for reroutable flow under complete disruption (Prop. 3 and 4, respectively). Optimal investment for nonreroutable flow can be characterized (Corollary 2 in Section 3.4.3)

4. In general, the problems of optimal investment for both routing mechanisms are NP-hard (Theorems 1 and 2 in Section 3.4.4). The expected max-flow for nonreroutable flow is a *monotone supermodular* set function in the sense of node investment (Theorem 3 in Section Proof 3.8).

   (a) A greedy algorithm (Algorithm 1 in Section Proof 3.9) can therefore facilitate practical decision-making of resilience investment for nonreroutable flow, with guaranteed performance.

We introduce the notation and functions below to facilitate the analysis.

- $\mathcal{L}$, the state space (i.e., support) for $\boldsymbol{L}$. $\mathcal{L} \subseteq \{1\} \times \{l, 1\}^{|V_S|} \times \{1\}$
- s, the scenario (i.e., realization) of the node states within the state space $\mathcal{L}$, i.e., $s \in \mathcal{L}$
- d_s, the number of disrupted nodes in the scenario s
- $\mathcal{D}_s$, the set of disrupted nodes in the scenario s. $\mathcal{D}_s \subseteq V_S$ and $|\mathcal{D}_s| = d_s$
- $\mathcal{L}_d$, the set of scenarios with d disruptions. $\mathcal{L}_d \subset \mathcal{L}$

Let $p^s(\mathcal{N})\colon 2^{V_S} \to \mathbb{R}_{\geq 0}$ be the function of the probability of the scenario s given investment $\mathcal{N}$.

$$p^s(\mathcal{N}) = \prod_{\substack{i\in\mathcal{D}_s\setminus\\(\mathcal{D}_s\cap\mathcal{N})}} p^D \prod_{i\in\mathcal{D}_s\cap\mathcal{N}} (p^D - \delta p^D) \prod_{\substack{i\in V_S\setminus\\(\mathcal{D}_s\cup\mathcal{N})}} (1-p^D) \prod_{\substack{i\in\mathcal{N}\setminus\\(\mathcal{D}_s\cap\mathcal{N})}} (1-p^D+\delta p^D), \quad \forall\mathcal{N}\subseteq V_S \tag{3.15}$$

We count p^D in Equation (3.15) and equivalently re-express $p^s(\mathcal{N})$ as $p(|\mathcal{N}|, d_s, u)$:

$$\begin{aligned} p^s(\mathcal{N}) &\equiv p(|\mathcal{N}|, d_s, u) \\ &= {p^D}^{d_s-u}(p^D - \delta p^D)^u (1-p^D)^{I-|\mathcal{N}|-d_s+u}(1-p^D+\delta p^D)^{|\mathcal{N}|-u}, \forall s, \forall\mathcal{N} \end{aligned} \tag{3.16}$$

where $u = |\mathcal{D}_s \cap \mathcal{N}| \in \mathbb{N}$ that satisfies $\max(0, |\mathcal{N}| + d_s - I) \leq u \leq \min(|\mathcal{N}|, d_s)$.

The expected max-flow can take the following generic form (superscript omitted):

$$\mathbb{E}h(\mathcal{N}) = \sum_{s\in\mathcal{L}} p^s(\mathcal{N})h_s = \sum_{d=0}^{I}\sum_{s\in\mathcal{L}_d} p^s(\mathcal{N})h_s, \quad \forall\mathcal{N}\subseteq V_S \tag{3.17}$$

where h_s denotes the generic realized max-flow under scenario s. Table 3.4 illustrates Equation (3.17).

3.4.1 Routing Flexibility

Routing flexibility (i.e., being reroutable) increases network resilience. We show in Lemma 1 that for any scenario, the reroutable mechanism generates a higher or equal max-flow than the nonreroutable mechanism. We then use Lemma 1 to show that reroutable flow yields a larger or equal maximized expected max-flow than non-reroutable flow (Proposition 1).

Lemma 1 For any scenario $s \in \mathcal{L}$, $h_s^R \geq h_s^N$.

Proof 3.1

Proof of Lemma 1. Let $\boldsymbol{\phi}_s^*$ be a vector of realized optimal path flows for $h_s^N, \forall s$ (multiple $\boldsymbol{\phi}_s^*$ may exist). Without loss of generality, the kth element in $\boldsymbol{\phi}_s^*$ is either ϕ_k^* or $\phi_k^* \cdot l$, determined by all $L_i, i \in k$ from s. For any $i \in V$, regardless of ϕ_k^* or $\phi_k^* \cdot l$, the kth element is smaller than or equal to $L_i \phi_k^*$. Hence, summing up elements in $\boldsymbol{\phi}_s^*$ with respect to i, we can obtain

$$\sum_{k \in \Phi | i \in k} [\phi_k^* \text{ or } \phi_k^* \cdot l] \leq \sum_{k | i \in k} L_i \phi_k^* = L_i \sum_{k | i \in k} \phi_k^* \leq L_i \omega_i, \quad \forall i \in V \tag{3.18}$$

where the second inequality holds due to Inequality (3.1) (implied by (3.11)). Hence, $\boldsymbol{\phi}_s^*$ satisfies constraint (3.14) and is feasible for Equation (3.13) that returns some $\tilde{h}_s^R = h_s^N$. Due to that h_s^R is the max-flow, $h_s^R \geq \tilde{h}_s^R = h_s^N, \forall s \in \mathcal{L}$. Hence, Lemma 1 holds. $\square$

Proposition 1 The maximized expected max-flow for reroutable flow is greater than or equal to that for nonreroutable flow, i.e., $y^R \geq y^N, \forall K \leq |V_S|$.

Proof 3.2

Proof of Proposition 1. Based on Lemma 1,

$$\mathbb{E}h^R(\mathcal{N}) = \sum_{s \in \mathcal{L}} p^s(\mathcal{N}) h_s^R \geq \sum_{s \in \mathcal{L}} p^s(\mathcal{N}) h_s^N = \mathbb{E}h^N(\mathcal{N}), \forall \mathcal{N} \subseteq V_S \tag{3.19}$$

Based on Equation (3.5), if there exists $\mathcal{N}^*$ that maximizes $\mathbb{E}h^N(\mathcal{N})$, then $y^R \geq \mathbb{E}h^R(\mathcal{N}^*) \geq \mathbb{E}h^N(\mathcal{N}^*) = y^N, \forall K \leq |V_S|$. Proposition 1 holds. $\square$

Proposition 1 corresponds well with Sheffi and Rice Jr (2005, p.41)'s statement that "resilience can be achieved by ... increasing flexibility". From a network perspective, being reroutable is one aspect of flexibility. We remark that being reroutable or nonreroutable is a mechanism or a design. In other words, the same network can have reroutable or nonreroutable flows, as depicted in Figures 3.2 and 3.3. Although the nonreroutable flow may be viewed as a special case of the reroutable flow, we do not suggest investigating only one, as both are contextually meaningful.

3.4.2 Rare Disruptions

Rare disruptions imply a small p^D. We investigate $p^D = o(1/I)$ so that any multiplication of two or more p^D tends to be 0 as $I \to \infty$. Based on Equation (3.15), under $p^D = o(1/I)$,

$$p^s(\mathcal{N}) \leq \prod_{\substack{i \in \mathcal{D}_s \setminus \\ (\mathcal{D}_s \cap \mathcal{N})}} p^D \prod_{i \in \mathcal{D}_s \cap \mathcal{N}} (p^D - \delta p^D) \leq \prod_{i \in \mathcal{D}_s} p^D \to 0 \text{ as } I \to \infty, \quad \forall s \in \mathcal{L}_{d \geq 2}, \forall \mathcal{N} \subseteq V_S \tag{3.20}$$

Hence, under rare disruptions with a sizable network,

$$\mathbb{E}h(\mathcal{N}) = \sum_{s \in \mathcal{L}_{d \leq 1}} p^s(\mathcal{N}) h_s = \sum_{s \in \mathcal{L}_0} p^s(\mathcal{N}) h_s + \sum_{s \in \mathcal{L}_1} p^s(\mathcal{N}) h_s, \quad \forall \mathcal{N} \subseteq V_S \tag{3.21}$$

Condition of $K = 1$.

First, we consider the resource limitation $K = 1$. The optimal strategy is to invest on the node with the largest (normal) capacity.

Proposition 2 Regardless of the routing mechanism, $\mathcal{N}^* = \{v_{(1)}\}$, for $K = 1$ and $p^D = o(1/I)$, where $\omega_{(1)} \geq \omega_{(2)} \geq \cdots \geq \omega_{(I)}$.

Proof 3.3

Proof of Proposition 2. Under $K = 1$, based on Equation (3.15), we have $p^s(\mathcal{N}) =$

$(1-p^D)^{I-1}(1-p^D+\delta p^D)$ and $h_s^N = h_s^R = 1, \forall s \in \mathcal{L}_0, \forall \mathcal{N} \subseteq V_S$. Hence,

$$\mathcal{N}^* = \underset{|\mathcal{N}|\leq K=1}{\arg\max} \sum_{s\in\mathcal{L}_1} p^s(\mathcal{N})h_s, \quad \forall \mathcal{N} \subseteq V_S \tag{3.22}$$

where $p^{s\in\mathcal{L}_1}(\mathcal{N})$ takes the form of $p(1,1,0)$ of $I-1$ times and $p(1,1,1)$ once.

We examine h_s. According to the set-up for node and edge capacities (Equation (3.3)),

$$h^N_{s|i \text{ disrupted}} = h^R_{s|i \text{ disrupted}} = 1-(1-l)\omega_i, \quad \forall i \in V_S, \forall s \in \mathcal{L}_1 \tag{3.23}$$

Hence, we can sort $\{h_{s\in\mathcal{L}_1}\}$ in the increasing order using the parenthesized subscripts, i.e., $h_{\mathcal{L}_1,(1)} \leq h_{\mathcal{L}_1,(2)} \leq \cdots \leq h_{\mathcal{L}_1,(I)}$, and obtain $h_{\mathcal{L}_1,(j)} = 1-(1-l)\omega_{(j)}, j = 1,\ldots,I$. Given $p(1,1,1) \leq p(1,1,0)$,

$$\sum_{s\in\mathcal{L}_1} p^s(\mathcal{N})h_s \leq p(1,1,1)\cdot h_{\mathcal{L}_1,(1)} + p(1,1,0)\cdot \sum_{j=2}^{I} h_{\mathcal{L}_1,(j)} \tag{3.24}$$

where the equality is feasible when $\mathcal{N} = \{v_{(1)}\}$, which is therefore the optimal investment strategy $\mathcal{N}^*$. Hence, Proposition 2 holds. In addition, if multiple nodes share the largest normal capacity, optimal investment may target any of those nodes. □

Condition of $K \geq 2$.

Next, we consider $K \geq 2$ while still keeping $p^D = o(1/I)$. The optimal strategy is to invest on the nodes with the K largest (normal) capacities.

Corollary 1 Regardless of the routing mechanism, $\mathcal{N}^* = \{v_{(1)}, v_{(2)}, \ldots, v_{(K)}\}$, for $K \geq 2$ and $p^D = o(1/I)$, where $\omega_{(1)} \geq \omega_{(2)} \geq \cdots \geq \omega_{(I)}$.

Proof 3.4

Proof of Corollary 1. Following Equations (3.21), (3.22) and (3.23),

$$\mathcal{N}^* = \underset{|\mathcal{N}|\leq K}{\arg\max} \sum_{s\in\mathcal{L}_1} p^s(\mathcal{N})h_s, \quad \forall\mathcal{N}\subseteq V_S \tag{3.25}$$

where $p^{s\in\mathcal{L}_1}(\mathcal{N})$ takes the form of $p(K,1,0)$ of $I-K$ times and $p(K,1,1)$ of K times, and $h^N_{\mathcal{L}_1,(j)} = h^R_{\mathcal{L}_1,(j)} = 1-(1-l)\omega_{(j)}, j = 1,\ldots,I$. Similarly, given that $p(K,1,1) \leq p(K,1,0)$ and $|\mathcal{N}| \leq K$,

$$\sum_{s\in\mathcal{L}_1} p^s(\mathcal{N})h_s \leq p(K,1,1)\cdot\sum_{j=1}^{K} h_{\mathcal{L}_1,(j)} + p(K,1,0)\cdot\sum_{j=K+1}^{I} h_{\mathcal{L}_1,(j)} \tag{3.26}$$

In other words, $\sum_{s\in\mathcal{L}_1} p^s(\mathcal{N})h_s$ achieves its maximum at $\mathcal{N}^* = \{v_{(1)},\ldots,v_{(K)}\}$. Hence, Corollary 1 holds. In addition, if there are multiple sets of K nodes with K largest ω's, any set is optimal. $\square$

3.4.3 Frequent Disruptions

Another extreme case is that every node is highly likely disrupted (known as "frequent disruptions"). Under frequent disruptions, we characterize that $p^D = 1-\varepsilon \to 1$ as $\varepsilon \to 0$. In this sense, the probability of a scenario in which non-invested nodes are healthy can be regarded as 0. Hence,

$$\begin{aligned}\mathbb{E}h(\mathcal{N}) = \sum_{s\in\mathcal{L}} p^s(\mathcal{N})h_s &= (1-\delta)^{|\mathcal{N}|}h_\emptyset + \sum_{i_1\in\mathcal{N}}(1-\delta)^{|\mathcal{N}|-1}\delta h_{i_1} \\ &+ \sum_{i_1,i_2\in\mathcal{N}}(1-\delta)^{|\mathcal{N}|-2}\delta^2 h_{i_1,i_2} + \cdots + \delta^{|\mathcal{N}|}h_{\mathcal{N}}\end{aligned} \tag{3.27}$$

where $h_{\subseteq\mathcal{N}}$ is the max-flow when all nodes are disrupted except some subset of $\mathcal{N}$.

It is easy to show that $h^N_\emptyset = h^R_\emptyset = l < 1$ and $(1-\delta)^{|\mathcal{N}|} + \sum_{i_1\in\mathcal{N}}(1-\delta)^{|\mathcal{N}|-1}\delta +$

$\sum_{i_1,i_2 \in \mathcal{N}}(1-\delta)^{|\mathcal{N}|-2}\delta^2 + \cdots + \delta^{|\mathcal{N}|} = 1$. Therefore, as long as we can find such an $\mathcal{N}$ that at least one $h_{\subseteq\mathcal{N}} > l$, the expected max-flow can increase over l. In the following analysis, $h_{\mathcal{N}}$ is examined due to $h_{\mathcal{N}} \geq h_{\subseteq\mathcal{N}}$.

Proposition 3 Under $p^D = 1 - \varepsilon \to 1$ as $\varepsilon \to 0$, $\mathbb{E}h^N(\mathcal{N})$ increases if and only if $\mathcal{N}$ entails at least one path.

Proof 3.5

Proof of Proposition 3. Equivalently, we prove that $h^N_{\mathcal{N}} > l$ if and only if $\mathcal{N}$ entails at least one path. Let $\boldsymbol{\phi}$ be *any* realization of feasible path flows for $h^N_{\mathcal{L}_0}$ (there may be more than one $\boldsymbol{\phi}$). For nonreroutable flow, when any one node on path k is disrupted, ϕ_k reduces to $l\phi_k$. Suppose the existence of a path $\tilde{k}$ entailed by $\mathcal{N}$. Based on the definition of $\boldsymbol{\phi}$, its element $\phi_{\tilde{k}}$ measures the path flow for $\tilde{k}$ when no node on $\tilde{k}$ is disrupted. Therefore,

$$h^N_{\emptyset} = l = l \sum_{k \in \Phi, k \neq \tilde{k}} \phi_k + l \cdot \phi_{\tilde{k}} < l \sum_{k \in \Phi, k \neq \tilde{k}} \phi_k + \phi_{\tilde{k}} \leq h^N_{\mathcal{N}} \tag{3.28}$$

where $l\phi_{\tilde{k}}$ is the path flow when at least one node on $\tilde{k}$ is disrupted. The sufficient condition holds.

We prove the necessary condition through contradiction. Suppose $h^N_{\mathcal{N}} > l$ but $\mathcal{N}$ does not entail a path (hence, every path has at least one disrupted node). In this sense, every path flow is ϕ_k times l for nonreroutable flow by definition, which indicates that the sum of path flows equals to l. In other words, $h^N_{\mathcal{N}} = h^N_{\emptyset} = l$, contradictory to the assumption. In general, Proposition 3 holds. $\square$

Investing on paths is sufficient to improve $\mathbb{E}h^R(\mathcal{N})$ for reroutable flow, but is not a necessary condition. In other words, if $\mathbb{E}h^R(\mathcal{N})$ is improved over l, it may be due to the benefit of routing flexibility and investment on some (but not all) nodes on the path. Nonetheless, in an extreme case of complete destruction ($l = 0$), path investment is necessary and sufficient to improve $\mathbb{E}h^R(\mathcal{N})$.

Proposition 4 Under $p^D = 1 - \varepsilon \to 1$ as $\varepsilon \to 0$ and $l = 0$, $\mathbb{E}h^R(\mathcal{N})$ increases if and only if $\mathcal{N}$ entails at least one path.

Proof 3.6

Proof of Proposition 4. According to Lemma 1, $h^R_{\mathcal{N}} \geq h^N_{\mathcal{N}}$. If $\mathcal{N}$ entails at least one path, according to Proposition 3, $h^R_{\mathcal{N}} \geq h^N_{\mathcal{N}} > l$. The sufficient condition holds.

The necessary condition relies on $l = 0$, which indicates that if any node on the path is disrupted, the path flow becomes exactly 0. The benefit of routing flexibility is eliminated. Hence, following the same contradiction logic of the proof of Proposition 3, $h^R_{\mathcal{N}} > l = 0$ only if $\mathcal{N}$ entails at least one path. Given the necessary and sufficient conditions, Proposition 4 holds. $\square$

At last, we consider optimal paths when there are multiple ways to invest in two or more paths. The optimal strategy is characterized in Corollary 2 based on Equation (3.10), without proof.

Corollary 2 Under $p^D = 1 - \varepsilon \to 1$ as $\varepsilon \to 0$, the optimal resilience investment for nonreroutable flow is to invest in paths in a subset $\mathcal{K}$ of Φ that satisfy

$$\mathcal{N}^* = \underset{\mathcal{K} \in \mathcal{P}(\Phi)}{\arg\max} \quad \sum_{k \in \mathcal{K}} \phi_k \cdot \delta^{|k|} \tag{3.29}$$

$$\text{s.t.} \quad \sum_{k \in \Phi | (i,j) \in k} \phi_k \leq w_{ij}, \quad \boldsymbol{\phi} \in \mathbb{R}^{|\Phi|}_{\geq 0}, \forall (i,j) \in E \tag{3.30}$$

$$\sum_{k \in \mathcal{K}} |k| \leq K \tag{3.31}$$

where $\mathcal{P}(\Phi)$ (the power set) is the set of all subsets of Φ, and $|k| \equiv |\{i \in k\}| - 2$ is the number of nodes on k excluding v_z and v_0.

In general, based on Equations (3.10) and (3.29) for nonreroutable flow, the focus of optimal investment includes (1) the path flows ϕ_k through the node (essentially the node capacity that bounds the sum of path flows), and (2) the probability of a healthy path, further captured by the path length $|k|$ and the disruption profile p^D and δ. Given the disruption profile, optimal investment tends to favor nodes with shorter (average) path lengths and higher path flows (higher node capacities).

3.4.4 General Probability of Disruption

This subsection extends the preceding analysis to a general probability of disruption p^D. We first show the hardness of solving the optimal investment problem for both routing mechanisms. As both problems are NP-hard, we characterize the modularity of the expected max-flow (a set function) to seek the opportunity of the near-optimal yet efficient algorithm. Based on the supermodularity of the expected nonreroutable max-flow, at last, we assert that the greedy algorithm can be used in reality to facilitate decision-making with a guaranteed performance.

Hardness of Optimal Investment.

This part shows that both problems of optimal resilience investment with respect to two routing mechanisms are NP-hard, which prompts us to seek near-optimal yet practically efficient algorithms in the following subsections. Intuitively, from the example (Table 3.4) that shows an exponential growth of state and search space with the number of nodes, a straightforward computation of expected max-flow is intractable. Theorems 1 and 2 formally characterize the complexity of the problems.

Theorem 1 The problem of optimal resilience investment defined as Model 1 for nonreroutable flow is NP-hard.

Proof 3.7

Proof of Theorem 1. We prove by showing that Model 1 is at least as hard as the classic 0-1 knapsack problem, which has been shown NP-complete (Karp, 1972). Notice that Model 1 is a one-stage combinatorial optimization that maximizes over a binary series $\boldsymbol{x}$ and a nonnegative vector $\boldsymbol{\phi}$. Suppose we have found and then supply a feasible $\boldsymbol{\phi} \in \mathbb{R}_{\geq 0}^{|\Phi|}$ that satisfies constraint (3.11). In this sense, Model 1 reduces to a non-linear variant of the knapsack problem where the objective (3.10) involves a multiplication of the binary variables (i.e., $\prod x_i$).

With the 0-1 knapsack problem to be NP-complete, its non-linear variant (i.e., binary non-linear programming) is at least NP-complete. Moreover, with $\boldsymbol{\phi}$ to be determined additionally, the decision problem for Model 1 is at least NP-complete, and the optimization problem for Model 1 is NP-hard, whose resolution is at least as difficult as the decision problem and there is no known polynomial algorithm which can tell, given a solution, whether it is optimal. $\square$

Theorem 2 The problem of optimal resilience investment defined as Model 2 for reroutable flow is NP-hard.

Proof 3.8

Proof of Theorem 2. Although the maximum-flow problem in deterministic networks is solvable in polynomial time (Goldberg and Tarjan, 2014), its stochastic variant, i.e., obtaining the expected max-flow in a network that is subject to random edge failures (e.g., edge capacities follow a two-point distribution), is NP-hard (Ball, 1986; Nagamochi and Ibaraki, 1991).

Model 2 can be equivalently transformed to the expected max-flow problem as discussed in Nagamochi and Ibaraki (1991). Consider a new network $G' = (V', E')$, where we split every $i \in V_S$ into i_i and i_o in V'. V' has v_0 and v_z as well. For any $i \in V_S$, the edges in the form of (v_z, i), (k, i), (i, j), and (i, v_0) in E become (v_z, i_i), (k_o, i_i), (i_o, j_i), and (i_o, v_0) in E', respectively. Those kinds of edges in E' are assigned sufficiently large capacities and disruption probability 0. In addition, we add to E' edges (i_i, i_o) that have the same capacity and disruption probability as $i \in V_S$. In this sense, G' is equivalent to G but the max-flow is subject to the edge capacity constraint. Hence, Model 2 is at least as difficult as the expected max-flow problem as it searches over V_S to maximize the expected max-flow. The problem of optimal resilience investment is thus NP-hard. $\square$

Modularity of Expected Max-Flow.

We define a *monotone supermodular* set function and present Theorem 3 that characterizes the modularity of $\mathbb{E}h^N(\mathcal{N})$.

Definition 1 Let N be a finite ground set and $f\colon 2^N \to \mathbb{R}$. Then f is *supermodular* if for all $A \subseteq B \subseteq N$ and $i \in N \backslash B$,

$$f(A \cup \{i\}) - f(A) \leq f(B \cup \{i\}) - f(B) \tag{3.32}$$

Furthermore, a supermodular function f is *monotone* if $f(A) \leq f(B)$ when $A \subseteq B$.

Theorem 3 The expected max-flow for nonreroutable flow $\mathbb{E}h^N(\mathcal{N})$ as defined in Model 1 is a *monotone supermodular* function.

Proof 3.9

Proof of Theorem 3. We start by computing $\mathbb{E}h^N(A \cup \{i\}) - \mathbb{E}h^N(A)$ for all $A \subseteq B \subset$

V_S and $i \in V_S \backslash B$. Based on Equations (3.9) and (3.10), we can obtain

$$\mathbb{E}h^N(A \cup \{i\}) = l + (1-l)\sum_{k\in\Phi}\phi_k[\prod_{j\in k, j\neq i}(1-p^D+p^D\delta x_j)\cdot(1-p^D+\delta p^D)] \tag{3.33}$$

$$\mathbb{E}h^N(A \cup \{i\}) - \mathbb{E}h^N(A) = \delta p^D(1-l)\sum_{k\in\Phi|i\in k}\phi_k \cdot \prod_{j\in k, j\neq i}(1-p^D+p^D\delta x_j) \tag{3.34}$$

Comparing $\mathbb{E}h^N(A \cup \{i\}) - \mathbb{E}h^N(A)$ to $\mathbb{E}h^N(B \cup \{i\}) - \mathbb{E}h^N(B)$, the difference lies only in the part of $\prod_{j\in k, j\neq i}(1-p^D+p^D\delta x_j)$. For any $k \in \Phi$, the investment B implies a higher number of x_j's to be 1, hence a higher $\prod_{j\in k, j\neq i}(1-p^D+p^D\delta x_j)$ than investment A ($\delta > 0$). Hence, $\mathbb{E}h^N(A \cup \{i\}) - \mathbb{E}h^N(A) \leq \mathbb{E}h^N(B \cup \{i\}) - \mathbb{E}h^N(B)$, indicating that $\mathbb{E}h^N(\mathcal{N})$ is supermodular.

On the other hand, per Equation (3.17), $p^s(\mathcal{N})$ is non-decreasing with $\mathcal{N} \subseteq V_S$ and all h_s^N's are non-negative. Hence, $\mathbb{E}h^N(\mathcal{N})$ is monotone. □

However, the expected max-flow for reroutable flow is *neither* supermodular *nor* submodular (a function f is submodular if $-f$ is supermodular), which is easy to verify through the foregoing simple network example (refer to Figure 3.1).

Greedy Algorithm with Guaranteed Performance.

To facilitate the practical decision-making of resilience investment, we base on Bai and Bilmes (2018) and assert that the greedy algorithm (Algorithm 1) is still good to maximize $\mathbb{E}h^N(\mathcal{N})$. For cardinality-constrained maximization of the monotone supermodular function $\mathbb{E}h^N(\mathcal{N})$, the greedy algorithm is guaranteed to obtain a solution $\hat{\mathcal{N}}$ such that

$$\frac{\mathbb{E}h^N(\hat{\mathcal{N}})}{\mathbb{E}h^N(\mathcal{N}^*)} \geq 1-\kappa \tag{3.35}$$

where $\mathcal{N}^* \in \arg\max_{|\mathcal{N}|\leq K} \mathbb{E}h^N(\mathcal{N})$ and κ is the *supermodular curvature* as:

$$\kappa = 1 - \min_{v\in V_S} \frac{\mathbb{E}h^N(v|\emptyset)}{\mathbb{E}h^N(v|V_S\backslash\{v\})} \tag{3.36}$$

where $\mathbb{E}h^N(v|A)$ denotes $\mathbb{E}h^N(A\cup\{v\}) - \mathbb{E}h^N(A)$. Based on Equation (3.34), we can obtain

$$\frac{\mathbb{E}h^N(v|\emptyset)}{\mathbb{E}h^N(v|V_S\backslash\{v\})} = \frac{\sum_{k|v\in k}\phi_k \prod_{i\in k, i\neq v}(1-p^D)}{\sum_{k|v\in k}\phi_k \prod_{i\in k, i\neq v}(1-p^D+\delta p^D)} \tag{3.37}$$

Hence, when disruptions are rare ($p^D \to 0$), the minimum of Equation (3.37) tends to be 1, indicating that the greedy algorithm tends to solve the maximization problem. On the contrary, when disruptions are frequent ($p^D \to 1$), the minimum of Equation (3.37) tends to be 0, indicating that the greedy algorithm may not be theoretically guaranteed to maximize $\mathbb{E}h^N(\mathcal{N})$. In general, the rarer disruptions, the more effective the greedy algorithm is in maximizing the expected max-flow.

Algorithm 1 Greedy algorithm for resilience investment for nonreroutable flow

Input: $\mathbb{E}h^N(\cdot)$, G, and K.
Output: An approximation solution $\hat{\mathcal{N}}$.
Initialize: $\mathcal{N}^0 \leftarrow \emptyset$, $i \leftarrow 0$, and $R \leftarrow V_S$
while $\exists v \in R$, s.t. $|\mathcal{N}^i \cup \{v\}| \leq K$ **do**
  $v \in \arg\max_{v\in R} \mathbb{E}h^N(v|\mathcal{N}_i)$. {If more than one element, pick any one as v}
  $\mathcal{N}^{i+1} \leftarrow \mathcal{N}^i \cup \{v\}$.
  $R \leftarrow R\backslash\{v\}$.
  $i = i + 1$.
end while
Return $\hat{\mathcal{N}} \leftarrow \mathcal{N}^i$.

3.5 Data-Driven Prescription on Resilience Investment

The previous section focuses on (1) extreme disruption probabilities and (2) general disruption for nonreroutable flow, which are analytically tractable. Node and path investment strategies are examined for the extreme disruption cases. However, the investment decisions under mid-level disruptions and/or for reroutable flow remain unexplored, due to analytical intractability. In other words, with the goal to maximize the expected max-flow, it is so far not clear where to invest under mid-level disruptions and/or for reroutable flow. Hence, we adopt the data-driven approach (Simchi-Levi, 2013). We use the data that comes from the optimal investment of a realistic network and verify the driving forces. Specifically, we examine the interaction between node characteristics and exogenous contextual factors on the investment focus between node versus path.

The realistic network under investigation is a distribution network from Company A (actual name withheld for confidentiality reasons), a leading supply chain management company. As a robustness check, we examine two networks derived from Company A's network (see Appendix A.1).

3.5.1 Context and the Distribution Network

Company A's distribution network distributes packages for a retailer. The network has distribution centers (DCs) as nodes as well as inter-DC shipping routes as edges. DCs have capacities to sort and distribute packages. Disruptions (e.g., sudden labor shortage or machine breakdown) can reduce DC's capacity and cause package delays for consumers and losses for Company A. Hence, Company A needs to allocate limited resources to improve the resilience of the network to maintain output (i.e., the daily processed packages) in the face of a disruption.

We first construct a network map of Company A's supply chain network. In their network we use the DC with the highest degree centrality as the source node[1] and create a virtual sink. We map the logistics as edges and direct edges from the source to the sink node. Nodes with the same distance from the source do not have large amounts of flow between each other. Corresponding edges are removed to simplify the analysis.

Utilizing Company A's rich data, we determine the node and edge capacities from the number of packages going through every DC and every route for the year 2017. We normalize node capacities and let the source have capacity 1. Edge capacities are normalized according to the emanating node's capacity, and node capacities are determined by the sum of incoming edge capacities.

Figure 3.4 displays the final distribution network ("DN") in the tree layout. The source is at the top, while the virtual sink is not shown. The final network contains 45 investable DCs and 86 edges (not including virtual edges to the sink). Nodes are colored according to their normalized capacities (we take a square-root for more clear visualization). Table 3.5 presents node information (summary statistics for node capacity: max: 0.221, min: 2.24e-5, mean: 0.044, median: 0.025, standard deviation: 0.056).

3.5.2 Analysis and Results

In the analysis, we set the effect of a disruption to $l = 20\%$ for every node, without loss of generality. The optimal investment that maximizes the expected max-flow is examined in the following 9 conditions: (1) fixed absolute benefit (δp^D) at 0.1 and

[1]Through empirical examination, the selected DC has the highest demand within the study period. We can make a fair assumption that packages arrive in the selected DC first and are then distributed.

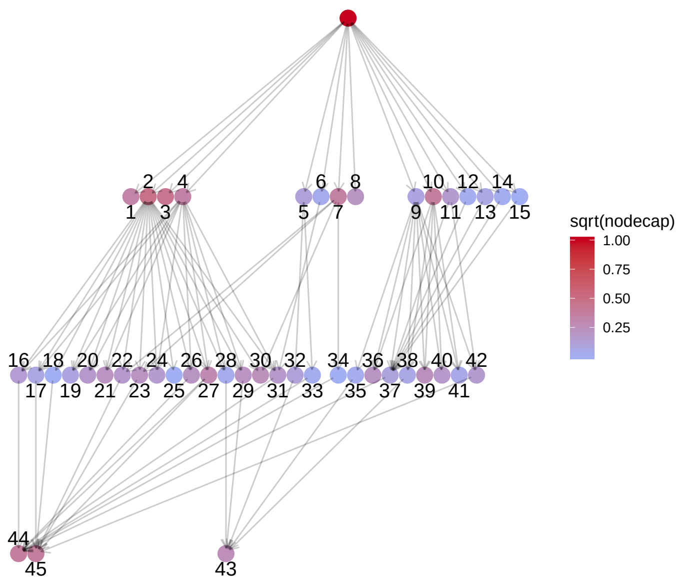


Figure 3.4: The final distribution network (DN) in the tree layout

Table 3.5: Node capacity with rank (in brackets) and # paths through node for DN

Node	Capacity	$\#\phi$	Node	Capacity	$\#\phi$	Node	Capacity	$\#\phi$
1 [8]	0.112	1	16 [26]	0.0204	2	31 [19]	0.0307	3
2 [1]	0.221	17	17 [32]	5.99e-03	2	32 [28]	0.0128	1
3 [2]	0.197	1	18 [44]	2.80e-05	1	33 [39]	2.46e-04	1
4 [7]	0.116	12	19 [30]	6.90e-03	2	34 [45]	2.24e-05	1
5 [27]	0.0131	2	20 [21]	0.0295	2	35 [38]	5.37e-04	1
6 [36]	1.04e-03	1	21 [14]	0.0453	2	36 [16]	0.0434	2
7 [6]	0.127	4	22 [22]	0.0274	2	37 [29]	9.76e-03	6
8 [18]	0.0418	1	23 [13]	0.0462	2	38 [34]	3.48e-03	2
9 [31]	6.48e-03	8	24 [23]	0.025	2	39 [11]	0.0585	2
10 [5]	0.138	5	25 [43]	4.31e-05	1	40 [20]	0.0306	2
11 [24]	0.0228	2	26 [17]	0.0422	2	41 [35]	2.91e-03	2
12 [40]	2.29e-04	1	27 [9]	0.0815	4	42 [25]	0.0228	2
13 [33]	3.99e-03	1	28 [37]	1.02e-03	2	43 [10]	0.0622	8
14 [41]	1.86e-04	1	29 [15]	0.0444	2	44 [4]	0.143	17
15 [42]	4.58e-05	1	30 [12]	0.0579	2	45 [3]	0.144	11
Total Paths: 58						Grand Mean:	0.0444	3.311

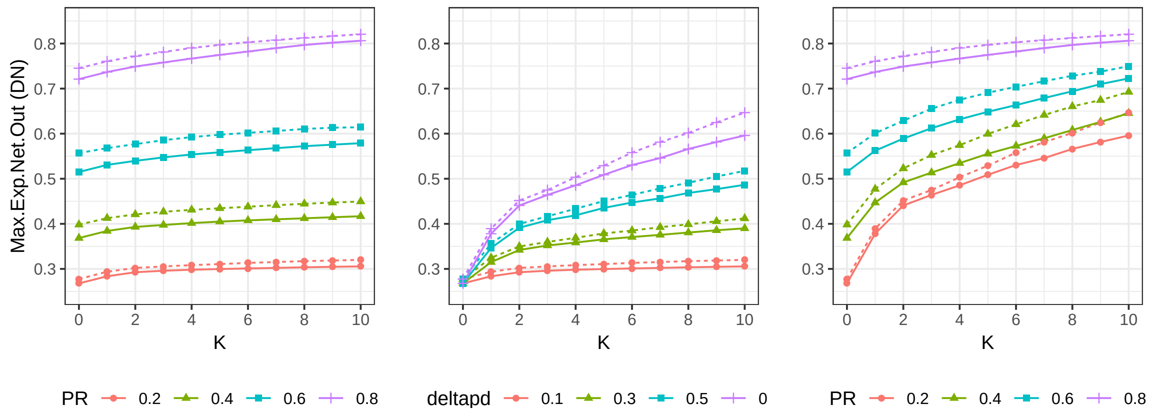


Figure 3.5: Max expected max-flow fixing absolute benefit (left), p^R (middle), and investment goal (right)

$p^R \equiv 1 - p^D$ at $0.2, 0.4, 0.6, 0.8$; (2) fixed p^R at 0.2 and absolute benefit (δp^D) at $0.1, 0.3, 0.5, 0.7$; and (3) fixed investment goal where $p^R = 0.2, 0.4, 0.6, 0.8$ and absolute benefit (δp^D) at $0.7, 0.5, 0.3, 0.1$. As the purpose and scope of this study lies in data-driven managerial prescription, we rely on extant techniques to obtain the optimal investment. Specifically, for nonreroutable flow, we solve Model 1 directly for optimal prescription. For reroutable flow, we adapt techniques designed for maximizing stochastic network flows, including simulated annealing in Janjarassuk and Nakrachata-Amon (2015), a sequential approximation algorithm in Cormican et al. (1998), and the WARMSTART algorithm in Sharma and Ghosh (2013).

The three techniques return consistent investment results for DN for each condition and routing (Table 3.6 displays the investment for DN). Hence, we conclude the optimality of the results, obtained although through simulation or near-optimal techniques. Figure 3.5 displays the trajectories of the maximized expected max-flow for DN as K increases from 0 to 10. As the dashed curves lie above their solid counterparts in the plots, routing flexibility is visually shown beneficial.

Table 3.6: Optimal investment for DN as K increases from 1 to 10

Different $(p^R, \delta p^D)$ combinations, under nonreroutable flow					
K	$(0.2, 0.1)$	$(0.4, 0.1)$	$(0.6, 0.1)$	$(0.8, 0.1)$	
1	3	3	3	3	
2	1,3	1,3	1,3	2,3	
3	1,3,8	1,3,10	1,2,3	1,2,3	
4	1,3,8,10	1,2,3,10	1,2,3,10	1,2,3,10	
5	1,2,3,8,10	1,2,3,8,10	1,2,3,7,10	1,2,3,10,44	
6	1,2,3,8,10,39	1,2,3,7,8,10	1,2,3,7,10,45	1,2,3,10,44,45	
7	1,2,3,7,8,10,39	1,2,3,7,8,10,39	1,2,3,7,10,44,45	1,2,3,7,10,44,45	
8	1,2,3,7,8,10,30,39	1,2,3,7,8,10,30,39	1,2,3,4,7,10,44,45	1,2,3,4,7,10,44,45	
9	1,2,3,7,8,10,30,36,39	1,2,3,4,7,8,10,30,39	1,2,3,4,7,8,10,44,45	1,2,3,4,7,10,27,44,45	
10	1,2,3,7,8,10,21,30,36,39	1,2,3,4,7,8,10,30,39,44	1,2,3,4,7,8,10,39,44,45	1,2,3,4,7,10,27,39,44,45	
K	$(0.2, 0.3)$	$(0.2, 0.5)$	$(0.2, 0.7)$	$(0.4, 0.5)$	$(0.6, 0.3)$
1	3	3	3	3	3
2	1,3	1,3	1,3	1,3	1,3
3	1,3,8	1,3,8	1,3,8	1,3,10	1,2,3
4	1,3,8,10	1,3,8,10	1,3,10,39	1,3,10,39	1,2,3,10
5	1,3,8,10,39	1,3,8,10,39	1,3,8,10,39	1,2,3,10,39	1,2,3,10,44
6	1,3,8,10,36,39	1,3,8,10,36,39	1,3,8,10,36,39	1,2,3,10,39,44	1,2,3,7,10,45
7	1,3,7,8,10,30,39	1,3,7,8,10,30,39	1,3,8,10,36,39,40	1,2,3,10,27,39,44	1,2,3,7,10,44,45
8	1,3,7,8,10,30,36,39	1,3,7,8,10,30,36,39	1,3,7,8,10,30,36,39	1,2,3,7,10,30,39,45	1,2,3,4,7,10,44,45
9	1,2,3,7,8,10,30,36,39	1,3,7,8,10,30,36,39,40	1,3,7,8,10,30,36,39,40	1,2,3,4,10,27,39,44,45	1,2,3,4,7,10,27,44,45
10	1,2,3,7,8,10,21,30,36,39	1,2,3,7,8,10,21,30,36,39	1,3,7,8,10,23,30,36,39,45	1,2,3,7,10,27,30,39,44,45	1,2,3,4,7,10,27,39,44,45
Different $(p^R, \delta p^D)$ combinations, under reroutable flow					
K	$(0.2, 0.1)$	$(0.4, 0.1)$	$(0.6, 0.1)$	$(0.8, 0.1)$	
1	3	3	3	3	
2	1,3	1,3	1,3	2,3	
3	1,3,8	1,2,3	1,2,3	1,2,3	
4	1,3,4,8	1,2,3,10	1,2,3,10	1,2,3,10	
5	1,2,3,4,8	1,2,3,10,45	1,2,3,10,45	1,2,3,10,44	
6	1,2,3,4,8,30	1,2,3,4,10,45	1,2,3,10,44,45	1,2,3,10,44,45	
7	1,2,3,4,8,10,30	1,2,3,4,10,30,45	1,2,3,10,30,44,45	1,2,3,4,10,44,45	
8	1,2,3,4,8,10,21,30	1,2,3,4,8,10,30,45	1,2,3,8,10,30,44,45	1,2,3,4,10,27,44,45	
9	1,2,3,4,8,10,21,30,44	1,2,3,4,8,10,30,44,45	1,2,3,4,10,27,30,44,45	1,2,3,4,7,10,27,44,45	
10	1,2,3,4,8,10,21,30,39,44	1,2,3,4,8,10,27,30,44,45	1,2,3,4,8,10,27,30,44,45	1,2,3,4,7,10,27,39,44,45	
K	$(0.2, 0.3)$	$(0.2, 0.5)$	$(0.2, 0.7)$	$(0.4, 0.5)$	$(0.6, 0.3)$
1	3	3	3	3	3
2	1,3	1,3	1,3	1,3	1,3
3	1,3,8	1,3,8	1,3,8	1,2,3	1,2,3
4	1,2,3,8	1,2,3,30	1,2,3,30	1,2,3,44	1,2,3,10
5	1,2,3,8,30	1,2,3,8,30	1,2,3,27,44	1,2,3,27,44	1,2,3,10,44
6	1,2,3,8,10,30	1,2,3,8,21,30	1,2,3,27,30,44	1,2,3,10,27,44	1,2,3,10,27,44
7	1,2,3,4,8,21,30	1,2,3,8,27,30,45	1,2,3,8,27,30,44	1,2,3,10,27,39,44	1,2,3,10,27,39,44
8	1,2,3,4,8,10,21,30	1,2,3,8,21,27,30,45	1,2,3,10,27,30,39,44	1,2,3,10,27,30,39,44	1,2,3,10,27,39,44,45
9	1,2,3,4,8,21,27,30,45	1,2,3,8,10,27,30,39,45	1,2,3,8,10,27,30,39,44	1,2,3,8,10,27,30,39,44	1,2,3,10,27,30,39,44,45
10	1,2,3,4,8,10,21,27,30,45	1,2,3,8,10,21,27,30,39,45	1,2,3,8,10,27,30,36,39,44	1,2,3,8,10,27,30,36,39,44	1,2,3,4,10,27,30,39,44,45

Hypotheses and the Empirical Model.

The analytical analysis suggests that firms focus their investments on either the node or path strategy under extreme disruption probabilities. For general cases of disruption, based on analytical insights, we conjecture that node characteristics interact with external contextual factors which influence the optimal investment strategy. The following three node characteristics are of concern: capacity ω_i, average path length $\overline{|k|}_i$, and the *flow centrality* $C_\phi(i)$. We do not consider the path flow ϕ_k in the analysis because (1) path flow is constrained by node capacity for both routing, and (2) node capacity is more straightforward and commonly estimated than path flow in reality. We consider the flow centrality because nodes with higher $C_\phi(i)$ may draw more investment to augment the benefit of routing flexibility under reroutable flow. The *flow centrality* $C_\phi(i)$ is defined as the number of paths going through node i over the number of total paths in G, i.e.,

$$C_\phi(i) = \frac{|\{k \in \Phi | i \in k\}|}{|\Phi|} \tag{3.38}$$

Two external contextual factors are of concern as well: the disruption probability p^D and the routing mechanism.

Generally, the analytical insights indicate optimal investment on nodes with higher capacities and on shorter paths, regardless of disruption probability and routing mechanism. However, nodes with higher (lower) capacities may lie on longer (shorter) paths. In order to determine the contribution of ω_i, $\overline{|k|}_i$, and $C_\phi(i)$ to investment priority given various contextual factors, we adopt the data-driven approach and use the optimal investment results (Table 3.6) and node information (Table 3.5) as data input. Based on Equation (3.10) and the previous arguments on flow centrality, we propose the following hypotheses of the interaction effects for empirical examination.

Hypothesis 5 *Ceteris paribus*, increasing disruption probability reduces the adoption of node-capacity investment strategy for network resilience.

Hypothesis 6 *Ceteris paribus*, increasing disruption probability increases the adoption of average-path-length investment strategy for network resilience.

Hypothesis 7 *Ceteris paribus*, routing flexibility improves the network resilience through investing on nodes with higher flow centralities.

We use an ordered logistic regression to empirically verify the three hypotheses. The ordered logistic regression, also known as the ordered logit model, cumulative link model, or proportional odds model, is a commonly used method for ordinal response variables (McCullagh, 1980). This method estimates the maximum-likelihood using iteratively reweighted least squares. We record the *first appearance* of the nodes, up to $K = 10$, and use it as the response variable ("1" being chosen at $K = 1$ and hence most prioritized). Independent variables include the contextual factors (namely, p^D and "rout"), the node characteristics (namely, ω_i, $\overline{|k|}_i$, and $C_\phi(i)$), and their interactions. We include the investment benefit δ as a control variable. Equation (3.39) illustrates of the model with only the main effects. Other models with interaction terms follow similar expressions.

$$\begin{aligned}\lambda_j(\boldsymbol{x}) &= \ln \frac{\sum_{i=1}^{j} Pr(Y = i|\boldsymbol{x})}{\sum_{i=j+1}^{k} Pr(Y = i|\boldsymbol{x})} \\ &= \alpha_j + \beta_1 p^D + \beta_2 \delta + \beta_3 \text{rout} + \beta_4 \omega_i + \beta_5 \overline{|k|}_i \\ &\quad + \beta_6 C_\phi(i), \quad j = 1, \ldots, k-1\end{aligned} \tag{3.39}$$

where $\boldsymbol{x}$ represents the explanatory variables; Y is the *first appearance* and has k ordered categories; λ_j is the ratio of the probability of a response less than or equal to a given category $j = 1, 2, \ldots, k-1$ to the probability of a response greater than this category; α_j and β's are coefficients.

Data Analysis Results.

Table 3.7 gives the results of the ordered logistic regression results for DN. Model 2 is particular for nonreroutable flow (rout $= 0, N = 93$) while Model 3 is for reroutable flow (rout $= 1, N = 90$). The ordered logistic regression requires the proportional odds assumption, which is verified for all the models. We are more concerned with the direction and significance of the variables on which the prescription builds than the actual values of coefficients.

Model 1 indicates that node capacity, average path length, and flow centrality significantly affect the optimal investment. Without any interaction with external contextual factors, the effects of ω_i and $\overline{|k|}_i$ are as expected – optimal investment favors nodes with higher capacities and on shorter paths. Holding constant ω_i and $\overline{|k|}_i$, a lower $C_\phi(i)$ generally indicates a higher average path flow, hence contributing to higher investment priority.

Models 2 through 4 include the interaction terms between p^D and node capacity and average path length. For node capacity, its main effect is consistently significantly negative, but p^D weakens the main effect of node capacity by exhibiting positive significance. Hypothesis 5 is supported for both routing mechanisms. We interpret that less frequent disruptions (lower p^D) and higher node capacity contribute to higher investment priority (i.e., earlier appearance). For average path length, Models 2 and 4 have significant interactions, indicating that more frequent disruptions (higher p^D) and shorter average paths contribute to higher investment priority. The insignificant interaction in Model 3 may be due to that the optimal investment is not necessarily on lower $\overline{|k|}_i$ when p^D is higher under reroutable flow – the routing flexibility could result in alternative investments (e.g., nodes of high flow centrality). In general, the

Table 3.7: Ordered logistic regression results for DN

	Response: *First Appearance*				
	Model 1	Model 2	Model 3	Model 4	Model 5
p^D	0.329	-29.506^{***}	-13.316^{***}	-18.269^{***}	-18.669^{***}
	(0.700)	(5.196)	(3.616)	(2.879)	(2.896)
δ	-0.689	-0.990	-0.516	-0.616	-0.631
	(0.515)	(0.775)	(0.765)	(0.531)	(0.530)
rout	-0.147			-0.448	0.221
	(0.280)			(0.288)	(0.417)
ω_i	-44.231^{***}	-152.036^{***}	-88.019^{***}	-107.213^{***}	-109.168^{***}
	(4.708)	(22.448)	(15.777)	(12.480)	(12.616)
$\overline{\lvert k\rvert}_i$	3.275^{***}	0.180	1.471	1.079	1.032
	(0.393)	(1.103)	(0.957)	(0.700)	(0.703)
$C_\phi(i)$	5.372^{*}	20.389^{***}	6.672	10.811^{***}	14.409^{***}
	(2.718)	(4.657)	(4.247)	(2.906)	(3.385)
$p^D \times \omega_i$		111.491^{***}	62.924^{**}	77.136^{***}	79.310^{***}
		(26.206)	(19.535)	(15.062)	(15.193)
$p^D \times \overline{\lvert k\rvert}_i$		8.051^{***}	2.429	4.169^{***}	4.342^{***}
		(1.924)	(1.354)	(1.065)	(1.073)
rout $\times C_\phi(i)$					-6.343^{*}
					(2.865)
N	183	93	90	183	183
logLik	-304.360	-122.169	-145.641	-278.703	-276.226
AIC	638.72	276.34	323.28	591.41	588.45

Note: *p<0.05; **p<0.01; ***p<0.001

analysis supports Hypothesis 6.

Finally, we examine if flow centrality indeed plays a role in Model 5, which includes the interaction between routing and flow centrality. The negative significant interaction coefficient indicates that, compared to nonreroutable flow, reroutable flow values higher flow centrality more. Hypothesis 7 is supported. In other words, a node with higher flow centrality, although meaning a lower path flow through it on average, tends more to be invested under reroutable flow. For instance, node 4 in DN has a high flow centrality but relatively large average path length (2.667). It should be less favored under higher p^D. However, it appears earlier and more often in such conditions of higher p^D as $(0.2, 0.1)$, $(0.2, 0.3)$, and $(0.4, 0.1)$ in Table 3.6, under reroutable flow.

3.5.3 Extended Prescription for Special Network Configuration

We have considered generic supply chain networks so far. The data-driven prescription implies an optimal investment considering both node capacity and average path length under various levels of disruption frequencies. When disruptions are less frequent, nodes with higher capacities can be prioritized even if they may lie on longer paths; when disruptions are more frequent, shorter paths can be prioritized with relatively higher node capacities. We extend this optimal investment to networks of special configuration in this subsection, namely the "diamond-shaped" supply chain network, the scale-free network, and the small-world network.

First, the "diamond shape" is a commonly seen configuration of real supply chain networks (e.g., Ang et al., 2016; Wang et al., 2017). Applying our prescription, we suggest supply chain managers to protect "common suppliers", as they tend to have both high capacities and flow centralities. Even when common suppliers do not have top capacities (e.g., node 4 in DN), our findings still support the belief that they are worth the resilience investment, especially under reroutable flow.

Second, the scale-free network is another widely observed supply chain network configuration (e.g., Hearnshaw and Wilson, 2013; Sun and Wu, 2005). A "scale-free" supply chain network is efficient (Hearnshaw and Wilson, 2013) and exhibits (at least asymptotically) power-law degree distribution with a few "hub" nodes of high degree centralities and many "peripheral" nodes of low centralities. Similar as the diamond-shaped network, we prescribe to protect those "hub" nodes which are usually resourceful with social capitals (Borgatti et al., 2009). In physical distribution networks, "hub" nodes are often equipped with larger capacities to process gathering material flows from many "peripheral" nodes. As scale-free networks are vulnerable to targeted attacks (Albert et al., 2000), our prescription corresponds well to fortifying the vulnerable nodes (i.e., hubs) to best mitigate disruptions.

Finally, we apply our prescription to small-world networks that are "highly clustered like regular lattices, yet have small characteristic path lengths like random graphs" (Watts and Strogatz, 1998, p.441). Physical networks that exhibit the "small-world" properties include electrical power grids (Watts and Strogatz, 1998) and alternative food networks (Brinkley, 2018). Our prescription still applies by first identifying one of the nodes of top capacities as the source node. Since most nodes in small-world networks can reach others by small steps (shorter path length), we further identify (short) paths of high flows including the source. Dependent on the disruption frequency, we can invest resilience resources on a few or many identified paths, which essentially fortifies high-capacity *clusters* due to a high clustering coefficient of the small-world network.

3.6 Implications and Discussion

As supply chain networks become more complex, the decision of where to make investments to improve resilience becomes more complicated. However, data-driven prescriptive analytics can help make decisions about where to invest in the supply chain network. The findings from this study are easy to communicate and implement. We propose taking a holistic network perspective to determine where to invest in the supply chain network (i.e. which nodes to improve). Prior research did not take a network perspective, but instead relied on identifying critical nodes or paths in isolation of the overarching network when assessing network resilience. This research suggests that such solutions can lead to suboptimal results. Specifically, when deciding whether some nodes are worth the investment, managers should examine not only the node capacity (broadly speaking, node contribution to the max-flow), but also the paths associated with each node. Managers may reflect on their network's real operations and follow the principled guidelines below to make investment decisions.

First, managers can assess the disruption profiles of nodes and estimate the likelihood of a disruption in advance. Based on the probability of a disruption and the routing mechanism, managers can shift the investment focus between critical nodes versus paths accordingly. Managers can focus more on critical nodes for less frequent disruptions, while on critical paths for more frequent disruptions. Although, in general, a node with higher capacity and on a shorter path receives higher investment priority, the clarification of the driving forces of node-versus-path investment helps decisions in reality (i.e., nodes have lower capacities and/or are on longer paths).

Second, for the same supply chain network, the reroutable mechanism yields a more resilient network than the nonreroutable mechanism. A reroutable mechanism increases the flexibility in the supply chain network, and this research shows the benefits of such a mechanism. The results further show that nodes with higher centralities receive higher investment priority under the reroutable flow mechanism. In this sense, "common suppliers" or "hub nodes", which usually have high capacities and high flow centralities, are more important in determining where to invest in resilience under reroutable flows.

Third, the monotone supermodular property of the investment under nonreroutable flow validates the effectiveness of the greedy algorithm. In this sense, supply chain managers can follow the greedy policy and invest on important nodes without spending much effort or computation resources. As a result, the findings provide practical guidelines, especially for firms with limited computational resources or expertise.

This study is not without limitations. The stochastic network flow model by nature may not be applicable to the assembly system, where the missing of one component results in the incompleteness of a product and the addition of path flows does not make contextual sense. Although the normalized flow amount may represent the success probability, a different model can be necessary to capture the characteristics of an assembly system.

In addition, we have made a number of assumptions in the construction of our model that can be extended in future research. First, the supply chain network is directed acyclic, such that loops are not considered. Although this assumption has a realistic basis for some networks, future research may analyze a more general flow network that allows mutual transshipment between nodes.

Second, disruptions in the model are independent and random. Previous literature has investigated targeted attacks, and these might be fruitfully incorporated into future research. In addition, disruptions like natural disasters may affect facilities in a common geographic region. A more complicated mechanism of disruption (e.g., correlated or propagating failures) may be adopted to extend the current model.

Third, we assume the cost of investing resilience resources to be the same across different nodes. In reality, reaching and investing in remote supply chain nodes may be more difficult and costlier compared to neighboring nodes. A natural extension of this research would incorporate a tiered cost structure in which the investment cost increases with the node's distance to the sink.

Finally, and most importantly, Essay 2 is based on the assumption that a supply chain network is fully visible to its manager. Full visibility is relatively easy to achieve in networks like the distribution system. But for supply networks involving multiple, inter-dependent companies, partial visibility is perhaps more realistic. Such an analysis is outside the scope of this essay yet future researchers could examine, for example, the capacity game among actors under partial visibility, to advance the literature on supply chain network resilience.

Chapter 4

Supply Network Resilience Learning: A Data-Driven Exploratory Study

4.1 Introduction

Ford recently announced that they were halting production of their profitable F-150 trucks due to a fire at Meridian Magnesium, one of their major component suppliers. Meridian Magnesium has committed to rebuild their magnesium die-casting complex (the source of the problem) and learn from this disruption (Muller, 2018). A disruption at a supplier's facility can propagate to others in the network. The disruption at Meridian Magnesium propagated to Ford, and shut down the production of F-150 trucks. As a result, it is not enough for firms to manage disruptions within their own facilities (Hendricks and Singhal, 2005), but they also need to manage disruptions across their supply networks (Kim et al., 2015).

When suppliers experience a disruption, they seek to *learn* from the event and reduce the risk of future events (Fiksel et al., 2015). Organizations like Ford benefit when their supply networks learn from disruptions. *Supply network resilience learning* occurs when the supply network becomes more resilient due to individual *supplier* *learning* from disruptions. When suppliers learn from a disruption, they reduce the risk of future disruptions. Although organizational learning has been extensively studied, supply network learning and its relationship to supplier learning remains unexplored. Specifically, in the face of a disruption, suppliers can learn to prevent future disruptions (*learn to prevent*) and learn to better recover from future disruptions (*learn to recover*). Organizations like CISCO, for example, have invested extensive resources in mapping out their supply networks to help their suppliers learn from disruptions (Sáenz and Revilla, 2013). However, it is unclear how individual supplier learning translates into supply network learning to improve network resilience. This research takes a data-driven approach to investigate the following question: *How does the supply network learn from suppliers' disruptions?*

To address this question, we examine disruptions at both the individual supplier level (node) and the overall supply network level (ego network). For instance, Meridian Magnesium would be at the node level while Ford's supply base would be the ego network level. Kim et al. (2015) developed a metric of resilience at the supply ego network level. Drawing on their work, we use the *proportion of disrupted paths* (PDP) over time to measure supply network resilience learning. Supply network resilience learning occurs when the PDP reduces over time. An agent-based model helps understand how supplier learning at the node level influences learning at the supply network level, and it provides "a natural description of a system ... of autonomous decision-making agents" (Bonabeau, 2002, p.7280). In this setting, suppliers act as agents in the supply network where they face disruptions. After a disruption, the supplier learns to both prevent future disruptions and better recover from future disruptions. We empirically investigate the model using Honda's and Toyota's supply networks, and we derive random networks from Honda's supply network structure to expand our investigation.

The analysis takes a multi-method approach (Choi et al., 2016) and consists of three phases. First, an analytical model shows that, for a generic directed acyclical supply network, the suppliers' *learning to prevent* a disruption (versus *learning to recover*) has a stronger effect on supply network resilience learning. That is, regardless of other factors, a supply network becomes free of disruptions over an infinite time period if suppliers learn to prevent disruptions. However, since the average life span of a company is decreasing (Mochari, 2016), understanding the performance of a supply network over a finite time horizon becomes more relevant. Second, an agent-based simulation examines the effects of suppliers' learning-to-prevent and learning-to-recover over a finite time horizon. The agent-based model generates a large-scale dataset which shows that a supplier's learning-to-recover enhances network learning more when they face a lower risk of a disruption, while learning-to-prevent becomes more beneficial when the rate of diffusing a disruption is lower. That is, the effect of the suppliers' modes of learning on the supply network resilience learning depends on the risk characteristics. The third phase varies the learning rates across suppliers and shows that the more central suppliers' learning rates have a stronger effect on supply network resilience learning.

This study makes the following three contributions to the supply chain risk management and organizational learning literature. First, this is the first study, to the best of our knowledge, that examines the relationship between network-level resilience learning and node-level disruption risks. Classic learning curves have characterized firm (or node-level) learning as it relates to operational problems (Lapré et al., 2011). This study makes a multilevel connection between supplier learning from disruptions and network learning through an agent-based model and simulation. Second, the analysis shows the contingent role of the risk. These contingencies suggest propositions that offer strategic guidelines on how to more effectively improve supply network resilience. Finally, due to the lack of theory and the analytical difficulty in characterizing network-level learning, this research contributes to taking a data-driven approach to understand operational risks in the supply network. With limited data on supply disruption incidents, this study offers a systematic guideline to create, visualize, and analyze a large-scale simulated dataset.

The results from this study also have several managerial implications. First, the focal firm of the supply network can assess risks and make recommendations to aid suppliers' learning efforts. For instance, suppliers' learning-to-prevent and learning-to-recover may involve different strategies and resources. Understanding the contingencies for network learning helps the focal firm make effective use of their limited resources to improve network-level resilience. Second, central suppliers in the network play a more critical role in supply network resilience learning. The focal firm can strategically improve key suppliers' learning based on their positions in the network in order to improve the entire supply network's learning. In general, investing in suppliers' improvement requires considerable resources, knowing which suppliers and what mode of learning to focus on can help the focal firm make the best use of their limited resources.

The rest of Essay 3 is organized as follows. Section 4.2 gives a literature review on supply network resilience and learning in the supply network. Section 3.3 gives the formulation of the agent-based computational model. Section 4.4 analytically characterizes the infinite-time network resilience level, which demonstrates supply network resilience learning. Section 4.5 describes the experimental design for the agent-based simulation, which identifies the contingencies to improve network-level learning. Section 4.6 visualizes and analyzes the simulated data, and generates data-driven propositions and strategies. Section 4.7 discusses the findings, managerial implications, and future research.

4.2 Literature Review on Network Resilience and Learning

Increasingly, scholars recognize the importance of supply network resilience and that managing operational risks goes beyond the focal firm (Kim et al., 2015). A disruption at one supplier can propagate to another and wreak havoc on the entire supply network (Basole and Bellamy, 2014; Scheibe and Blackhurst, 2018). Individual suppliers need to prevent and recover from a disruption (Jüttner and Maklan, 2011; Ponomarov and Holcomb, 2009; Sheffi and Rice Jr, 2005). Some suppliers in the network may play a more critical role in the resilience of the overall network (Craighead et al., 2007). Nonetheless, the current literature on supply chain resilience tends to focus on either the supplier level (Pettit et al., 2013) or the network level (Nair and Vidal, 2011; Zhao et al., 2011). Research on the connection between suppliers' actions and supply network's resilience has been largely overlooked.

This study draws on the organizational learning literature to understand how supply network resilience improves. From this perspective, organizations learn from their disruption experiences to improve resilience. Organizational learning from experiences has received extensive attention in the literature (Argote, 2012), and has been characterized by two major learning curves (Lapré et al., 2011). One curve takes the following power form (e.g., Argote and Epple, 1990; Darr et al., 1995; Dutton and Thomas, 1984; Yelle, 1979):

$$c_q = c_1 q^{-b} \tag{4.1}$$

where c_q is the unit cost to produce the qth unit, c_1 is the unit cost to produce the first unit, and b is the learning rate. The power curve originally described how firms learn to reduce cost, but it has also been used to describe learning from diverse areas such as error reduction, failures, disaster likelihood, and accident rates. Scholars have used the power form to study risk mitigation in airline companies (Haunschild and Sullivan, 2002), US railroads (Baum and Dahlin, 2007), coal mining organizations (Madsen, 2009).

The other curve that characterizes learning takes the following exponential form (e.g., Lapré et al., 2000; Levy, 1965):

$$Q(q) = P[1 - e^{-(a+\mu q)}] \tag{4.2}$$

where $Q(q)$ is the rate of output after producing q units, P is a maximum output that a firm could potentially achieve, a represents the initial efficiency, and μ represents the rate of adaptation. Similar to the power curve, the exponential curve has also been applied to other metrics, such as the success rate, the recovery rate, and the production rate (Levy, 1965; Madsen and Desai, 2010; Norrman and Jansson, 2004).

Although learning at the organizational level has been well studied, learning at the (supply) network level is emerging. Previous studies in the learning literature have focused on the effects of population-level (network-level) factors on firm-level performance, such as the influence of supply network structure on firm innovation (Bellamy et al., 2014) and the impact of population-level actors on firm failure prevention (Madsen and Desai, 2018). However, the learning effect on the network-level performance metrics has not been examined. In practice, firms invest in their suppliers with an overarching aim of improving supply network resilience (Liker and Choi, 2004). Investing in network-level learning may well explain the sustained success of firms like Cisco and Toyota in the face of increasing levels of supplier risks.

Supply network learning may depend on two sources: (1) supplier-specific factors that affect supplier learning (see for example Argote (2012) and Lapré et al. (2011) for thorough review), and (2) network characteristics that affect supply network learning. Two supplier-specific factors, in particular, may influence network-level learning. First, individual suppliers' learning rates (i.e, rates of *learning-to-prevent* and *learning-to-recover*) affect network learning directly. Kim and Tomlin (2013) described a firm's resilience to disruptions in terms of prevention and recovery capacities. This aligns with our related concepts of learning-to-prevent and learning-to-recover, although they did not take a learning perspective. Second, the diffusion, disruption, and recovery probabilities of operational risks facing suppliers will affect suppliers' learning and supply network resilience (Tomlin, 2006). In terms of the network characteristic, suppliers that occupy more central positions in the network can potentially have a stronger effect on network learning. In general, a more central position gives a supplier more resources (Borgatti and Li, 2009) and importance to affect supply network resilience. Due to a lack of strong theory, this study takes a more data-driven approach to understand interactions between node-level and network-level learning.

4.3 Agent-Based Computational Model

This section formulates the computational model that characterizes the network's and the suppliers' (agents') behavior. The model formulation begins with a focal firm's supply network (the ego network). The supply network is a generic flow network with *paths* that represent the flow of physical materials. Let $G = (V, E)$ be a directed acyclic graph that represents the supply network, where V is the set of nodes and $E \subseteq V \times V$ is the set of edges that are ordered pairs of nodes. The direction of an edge is determined by the material flow from a supplier to a buyer. Let $v_0 \in V$ be the single sink in G, where v_0 is the focal firm or the ego of the supply network. To facilitate the analysis, we classify the nodes based on their "levels" in the acyclic structure. *Source nodes* do not have incoming edges and are at level 0. Nodes that have incoming edges *only* from level-0 nodes are at level 1. Nodes that have incoming edges *only* from level-0 and level-1 nodes (i.e., lower-level nodes) are at level 2, and so on, till there is only the sink node in the network. Let $V^n, 0 \leq n \leq M$ be the set of level-n nodes, where M denotes the largest level. Finally, a path is defined as a combination of nodes and edges emanating from a source node and terminating at v_0.

Every supplier is vulnerable to random disruption at discrete points in time. In addition, a disrupted supplier may recover at another time point. The supplier faces a risk or probability of disruption, diffusion, and recovery, which are denoted by $\beta \in (0, 1)$, $\alpha \in (0, 1)$, and $\gamma \in (0, 1)$, respectively. A supplier can be disrupted with probability β; this disruption can diffuse to neighboring suppliers with probability α, and a supplier can recover from a disruption with probability γ. Disruptions at suppliers can propagate or diffuse along their directed edges to other suppliers. The phenomenon of *risk propagation* is defined as "endogenous or exogenous risks propagate from one organization to other organizations through the supply network" (Basole and Bellamy, 2014, p.755), and has been examined by several scholars (e.g., Chatfield et al. (2013), Lee et al. (1997a), Wu et al. (2007)). A supplier can be disrupted either by its own risk or through risk propagation from its suppliers. For the parameters α, β, and γ, we exclude their values being 0 and 1 to avoid trivial cases or a lack of convergence in the model. This study focuses on the supply to the focal firm of the ego supply network. As a result, we assume the sink node is never disrupted. This assumption also aligns with the fact that major disruptions (85%) normally originate from outside of the focal firm (Company, 2018).

Each supplier follows a discrete-time stochastic process. A supplier i has two states, namely *disrupted* (D) and *healthy* (H) (see Figure 4.1). The state of a node at time $t + 1$ is determined by its state at time t and the transition probabilities β_i^t or γ_i^t at t.

The two transition probabilities of supplier i are affected by the characteristics

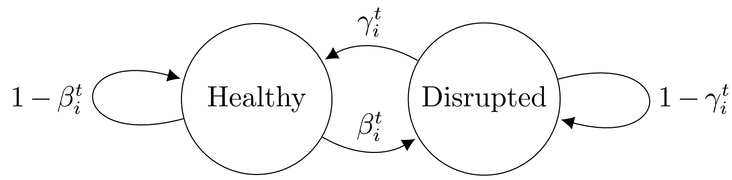


Figure 4.1: A stochastic process for node i at time t

of the risk and supplier i's two learning modes (prevent or recover). We make four assumptions about supplier learning. First, the supplier's two learning modes may involve different resources and strategies. Consequently, the rates of supplier i's learning-to-prevent and learning-to-recover may be different, denoted by a_i and b_i respectively. Second, we assume constant learning rates over time, although in reality the learning rates may vary over time (Lapré et al., 2011). Third, we assume suppliers can learn from all their prior disruption experiences, although in reality suppliers could potentially forget what they have learned (Agrawal and Muthulingam, 2015; Argote, 2012). Finally, we assume suppliers learn through their own disruption experiences rather than through the experiences of other organizations (Håkansson et al., 1999; Hora and Klassen, 2013).

Based on these assumptions, we characterize the effect of supplier learning on the transition probabilities. For β_i^t, supplier learning reduces the probability of a disruption (a negative outcome). Therefore, we adopt the power form of the learning curve as shown in Equation (4.1). For γ_i^t, supplier learning increases the probability of becoming healthy (a positive outcome). Hence, we adopt the exponential form of the learning curve as shown in Equation (4.2).

$$\beta_i^t = f(\beta[CumDis_i^t + 1]^{-a_i}) \tag{4.3}$$

$$\gamma_i^t = h(1 - (1 - \gamma)e^{-b_i CumDis_i^t}) \tag{4.4}$$

where f and h are two functions that involve risk impacts and are further characterized in Equations (4.6) and (4.7) respectively, and $CumDis_i^t$ is node i's cumulative disruption experiences from time 0 to t. We use $[CumDis_i^t+1]$ rather than $[CumDis_i^t]$ in Equation (4.3) to constrain $\beta[CumDis_i^t+1]^{-a_i}$ within $(0, \beta]$ and avoid a zero denominator. It is also easy to verify that $1-(1-\gamma)e^{-b_i CumDis_i^t} \in [\gamma, 1)$. The initial state for every supplier is determined by β.

Next, we characterize the effect of the risk on the transition probabilities. Although we could assign different initial disruption and recover probabilities across suppliers to reflect their heterogeneity, we make the initial disruption and recover probabilities the same across suppliers (β and γ in Equations (4.3) and (4.4)). We focus on risk propagation hereafter. Let m_i^t be the number of i's disrupted suppliers at time t. Hence, the probability that the risk does *not* diffuse to node i at time $t+1$ ("risk non-diffusion probability", or "RNDP") is:

$$\text{RNDP}_i^t \stackrel{\text{def}}{=} p(\text{Risk not diffused to } i \text{ from } t \text{ to } t+1) = (1-\alpha)^{m_i^t} \tag{4.5}$$

where risk diffusion activities are independent of each other. Noticeably, m_i^t is constrained by the supply network topology (i.e., i's in-degree centrality). In general, we characterize the transition probabilities of node i from time t to $t+1$ as

$$\beta_i^t = 1-(1-\beta[CumDis_i^t+1]^{-a_i}) \cdot (1-\alpha)^{m_i^t} \tag{4.6}$$

$$\gamma_i^t = [1-(1-\gamma)e^{-b_i CumDis_i^t}] \cdot (1-\alpha)^{m_i^t} \tag{4.7}$$

Following Kim et al. (2015), we use the *proportion of disrupted paths* (PDP) over time as the measure of network-level resilience. This metric considers a path disrupted at time t if there is at least one disrupted node on the path at t. PDP is computed as the ratio between the number of disrupted paths and the total number of paths in G. Table 4.1 summarizes the notation used in the model and the analysis.

Table 4.1: Notation in the model and analysis

Sets	Description
V	Set of nodes in the graph G, $V = \cup_{n=0}^{M} V^n \cup \{v_0\}$
V^n	Set of level-n nodes in G, $V^n \subset V, 0 \leq n \leq M$
E	Set of edges in the graph G
Φ	Set of paths in the graph G
Parameters	**Description**
a_i	Supplier i's rate of learning-to-prevent
b_i	Supplier i's rate of learning-to-recover
$CumDis_i^t$	Supplier i's cumulative disruptions till time t (inclusive)
D	Supplier's state of being disrupted
H	Supplier's state of being healthy
m_i^t	Number of node i's disrupted suppliers at time t
M	The largest number of node levels
RNDP_i^t	The probability of risk not diffused to i from time t to $t+1$
PDP^t	The proportion of disrupted paths in G at time t
t	The time index
α	The probability of diffusing a disruption
β	The probability of disruption
γ	The probability of recovery from a disruption
β_i^t	Supplier i's H-to-D transition probability for $t+1$'s state
γ_i^t	Supplier i's D-to-H transition probability for $t+1$'s state
π_i^H	Supplier i's limiting probability of the healthy state
π_i^D	Supplier i's limiting probability of the disrupted state
ϕ	A path in the graph G, $\phi \in \Phi$

4.4 Supply Network Resilience at Infinite Time

This section investigates supply network resilience learning for the infinite-time proportion of disrupted paths ($\text{PDP}^\infty \stackrel{\text{def}}{=} \lim_{t\to\infty} \text{PDP}^t$) based on the agent-based computational model. The analysis begins with the node-level stochastic processes, and then characterizes the network-level metric of resilience at infinite time (PDP^∞) based on the limiting distribution of the node states.

4.4.1 Limiting Distribution of Node States

The analysis begins with the source nodes. Since a source node of the ego network by definition does not have suppliers, its stochastic process is independent of other nodes. A source node $i \in V^0$ therefore has $m_i^t \equiv 0, \forall t$. Based on Equations (4.6) and (4.7), i's transition probabilities become

$$\beta_i^t = \beta[CumDis_i^t + 1]^{-a_i}, \quad \forall i \in V^0 \tag{4.8}$$

$$\gamma_i^t = 1 - (1-\gamma)e^{-b_i CumDis_i^t}, \quad \forall i \in V^0 \tag{4.9}$$

$CumDis_i^t$ is non-decreasing with time. As time increases, the frequency of disruption reduces while the chance of recovery improves. As time goes to infinity, we can show that

$$\lim_{t\to\infty} CumDis_i^t = \infty, \quad \forall i \in V^0 \tag{4.10}$$

through proof by contradiction via Equation (4.8). With infinite cumulative disruptions, depending on the source node i's learning modes (to prevent and/or to recover), i's limiting transition probabilities can be characterized as

$$\lim_{t\to\infty} \beta_i^t = \begin{cases} \beta[\lim_{t\to\infty} CumDis_i^t + 1]^{-a_i} = 0, & \text{for } a_i > 0, \forall i \in V^0 \\ \lim_{t\to\infty} \beta = \beta, & \text{for } a_i = 0, \forall i \in V^0 \end{cases} \tag{4.11}$$

$$\lim_{t\to\infty} \gamma_i^t = \begin{cases} 1 - (1-\gamma)e^{-b_i \lim_{t\to\infty} CumDis_i^t} = 1, & \text{for } b_i > 0, \forall i \in V^0 \\ \lim_{t\to\infty} \gamma = \gamma, & \text{for } b_i = 0, \forall i \in V^0 \end{cases} \tag{4.12}$$

Since the limiting transition probabilities are constant numbers for all situations of learning (Equations (4.11) and (4.12)), the limiting distribution exists for the source node.

$$\pi_i = [\pi_i^H, \pi_i^D] = \left[\frac{\lim_{t\to\infty}\gamma_i^t}{\lim_{t\to\infty}\beta_i^t + \lim_{t\to\infty}\gamma_i^t}, \quad \frac{\lim_{t\to\infty}\beta_i^t}{\lim_{t\to\infty}\beta_i^t + \lim_{t\to\infty}\gamma_i^t}\right]$$
$$= \begin{cases} [1, 0] & \text{for } a_i > 0, \forall i \in V^0 \\ [1/(\beta+1), \beta/(\beta+1)], & \text{for } a_i = 0, b_i > 0, \forall i \in V^0 \\ [\gamma/(\beta+\gamma), \beta/(\beta+\gamma)], & \text{for } a_i = 0, b_i = 0, \forall i \in V^0 \end{cases} \tag{4.13}$$

Lemma 2 shows a more stringent argument that the limiting distribution exists for any supplier in the network, regardless of their learning rates. See Appendix B.1 for the proof of Lemma 2.

Lemma 2 The limiting distribution exists for any non-sink node $i \in V\backslash\{v_0\}$, in the form of

$$\pi_i = [\pi_i^H, \pi_i^D] = \left[\frac{\lim_{t\to\infty}\gamma_i^t}{\lim_{t\to\infty}\beta_i^t + \lim_{t\to\infty}\gamma_i^t}, \quad \frac{\lim_{t\to\infty}\beta_i^t}{\lim_{t\to\infty}\beta_i^t + \lim_{t\to\infty}\gamma_i^t}\right]$$
$$= \begin{cases} \left[\mathrm{RNDP}_i^\infty, \quad 1 - \mathrm{RNDP}_i^\infty\right] & \text{for } a_i > 0, b_i > 0, \forall i \in V^n \\ \left[\frac{\gamma \cdot \mathrm{RNDP}_i^\infty}{1-(1-\gamma)\mathrm{RNDP}_i^\infty}, \quad \frac{1-\mathrm{RNDP}_i^\infty}{1-(1-\gamma)\mathrm{RNDP}_i^\infty}\right] & \\ & \text{for } a_i > 0, b_i = 0, \forall i \in V^n \\ \left[\frac{\mathrm{RNDP}_i^\infty}{1+\beta\cdot\mathrm{RNDP}_i^\infty}, \quad \frac{1-(1-\beta)\mathrm{RNDP}_i^\infty}{1+\beta\cdot\mathrm{RNDP}_i^\infty}\right] & \text{for } a_i = 0, b_i > 0, \forall i \in V^n \\ \left[\frac{\gamma\cdot\mathrm{RNDP}_i^\infty}{1-(1-\beta-\gamma)\mathrm{RNDP}_i^\infty}, \quad \frac{1-(1-\beta)\mathrm{RNDP}_i^\infty}{1-(1-\beta-\gamma)\mathrm{RNDP}_i^\infty}\right] & \\ & \text{for } a_i = 0, b_i = 0, \forall i \in V^n \end{cases} \tag{4.14}$$

where $\mathrm{RNDP}_i^\infty \stackrel{\text{def}}{=} \prod_{j|(j,i)\in E}(1-\alpha\pi_j^D)$ $(\mathrm{RNDP}_i^\infty \equiv 1, \forall i \in V^0)$ and $0 \leq n \leq M$.

Lemma 2 implies a special case of $\mathrm{RNDP}_i^\infty = 1$, where the limiting distribution of a supplier reduces to a simple form of Equation (4.13). In order to achieve $\mathrm{RNDP}_i^\infty = 1$, all of i's immediate suppliers should reach $\pi_j^D = 0, \forall j|(j,i) \in E$ with any non-trivial positive α. Since $1-(1-\beta)\mathrm{RNDP}_j$ is strictly positive, based on Equation (4.14),

j has to reach $\text{RNDP}_j = 1$ and learn to prevent as well. By continuing to derive this backwards, the entire supply base of i has to learn to prevent and maintain a zero limiting probability of disruption. In this sense, in order for a supplier i to have a limiting distribution in the form of Equation (4.13), all nodes in i's supply base should learn to prevent. Consequently, at infinite time, i's supply base converges to staying healthy.

For a sufficiently long but finite period of time, the long-term state distribution of a node is concentrated at state H if the entire network learns to prevent. Transitions to the state D become less and less frequent, till infinitely approaching while still larger than 0. Therefore, every supplier will approximately stay healthy with very sparse occurrence of the disruption.

4.4.2 Expectation of Infinite-Time PDP

With the existence of infinite-time limiting distribution for all the suppliers, the distribution of the PDP at infinite time can be determined. Proposition 5 characterizes the expectation of PDP^∞. To facilitate the discussion, we use $i \in \phi$ to denote that node $i \in V$ is on the path ϕ.

Proposition 5 The expectation of the infinite-time PDP is

$$\mathbb{E}[\text{PDP}^\infty] = 1 - |\Phi|^{-1} \sum_{\phi \in \Phi} \prod_{i \in \phi} \pi_i^H \tag{4.15}$$

where π_i^H is the limiting probability of i's healthy state from Equation (4.14).

See Appendix B.1 for the proof of Proposition 5. Proposition 5 suggests some special cases. First, when all suppliers have the same learning mode(s) (not necessarily same rates), there exists a lower bound for the expected PDP$^\infty$.

$$\mathbb{E}[\mathrm{PDP}^\infty] \geq 1 - |\Phi|^{-1} \sum_{\phi \in \Phi} (\pi^H_{V^0 \cap \phi})^{|\phi|-1} \tag{4.16}$$

where $V^0 \cap \phi$ stands for the source node on a path ϕ and $|\phi| - 1$ is the path length excluding the sink node. When all suppliers adopt the same learning mode(s), all source nodes then share the same $\pi^H_{V^0 \cap \phi}$ value (can be obtained from Equation (4.13)). Because the non-source node's limiting probability of state H is always smaller than or equal to the source node's limiting probability of state H (because $\mathrm{RNDP}_i^\infty \leq 1, \forall i$), π_i^H in Equation (4.15) satisfies $\pi_i^H \leq \pi^H_{V^0 \cap \phi}$, hence Equation (4.16).

Second, Equation (4.16) can be further reduced in a more special case where the supply network has paths of the same length.

$$\mathbb{E}[\mathrm{PDP}^\infty] \geq 1 - (\pi^H_{V^0 \cap \phi})^{|\phi|-1}, \quad \forall \phi \in \Phi \tag{4.17}$$

Third, when all suppliers learn to prevent, we can obtain $\pi_i^H = 1, \forall i \in V$ (based on Equations (4.13) and (4.14)). Hence, $\mathbb{E}[\mathrm{PDP}^\infty] = 0$. Notice that the PDP is non-negative. Therefore, the PDP$^\infty$ itself, not only the expectation, remains 0, regardless of the rate of suppliers' learning-to-prevent and whether some or all suppliers learn to recover (Corollary 3).

Corollary 3 PDP$^\infty$ $= 0$ when all suppliers learn to prevent (i.e., $a_i > 0, \forall i \in V \backslash \{v_0\}$).

The third special case exactly demonstrates the supply network resilience learning, as the disruption will eventually reduce to 0 in infinite time. In addition, although $\pi_i^H = 1$ seems a special case of Proposition 5, the behavior of the PDP is fundamentally different between $\pi_i^H = 1$ and $\pi_i^H \neq 1$. When all suppliers learn to prevent, the PDP reaches consensus and remains 0 at infinite time, while when some or no suppliers learn to prevent, the PDP keeps fluctuating even when the time goes to infinity, because node states vary (although in a steady manner), which makes the realized PDP fluctuate. Hence, $\pi_i^H = 1$ highlights the determinant role of suppliers' learning-to-prevent in keeping a constant infinite-time PDP (equal to 0) in the computational model.

4.5 Agent-Based Simulation and Experimental Design

Section 4.4 characterizes the behavior of the infinite-time proportion of disrupted paths and demonstrates supply network resilience learning from suppliers' disruptions. However, the infinite-time PDP itself cannot reflect the dynamic network learning process, i.e., the temporal trajectory of how the supply network learns to improve resilience. Given the analytical difficulty to characterize the finite-time PDP curve, we design two experiments that use the agent-based computational model to simulate PDP trajectories and characterize network-level learning.

4.5.1 Simulation and Network Setup

Scholars increasingly recognize the importance of simulation studies to understand complex emergent phenomena and to augment empirical data (Chandrasekaran et al., 2018). For example, simulation has been used to understand learning dynamics in the firm network (e.g., Baum et al., 2010) and risk propagation in the supply chain (e.g., Chatfield et al., 2013; Ivanov, 2017). In our case, supplier learning from disruptions is a complex process, and is further complicated through embeddedness in a supply network. We develop an agent-based simulation to understand network learning and explore contingencies that improve network learning.

We construct two realistic supply networks from Honda and Toyota using the Bloomberg SPLC database. The automobile supply networks exhibit tiered structures (Silver, 2016) that conform to our network model setup. Automobile supply networks are complex with tightly interconnected suppliers, which enables the propagation of operational risks (Simchi-Levi et al., 2015; Kim et al., 2011). In particular, Honda and Toyota are among the largest auto makers in the world, where they each have 9.2% and 14.7% of the global market respectively as of September 2018 (Statista, 2018). As a result, they provide good representatives of complex supply networks in the auto industry.

To develop Honda's and Toyota's supply networks, we search for their suppliers on the Bloomberg SPLC database, where a supplier has to account for more than 0.5% cost of goods sold of their immediate buyers. We restrict the number of tiers in the supply chain to 2, which conforms to "the visible bounded horizon of the focal agents" (Carter et al., 2015, p.93). The selected suppliers form the nodes of the network. Edges stand for the material flows and are formed according to the business relationships in the Bloomberg data. Honda's supply network has 96 nodes (including Honda as the focal firm) and 121 directed edges, while Toyota's supply network has 88 nodes (including Toyota) and 129 directed edges. Figure 4.2 visualizes both supply networks where the red node is the focal firm of the network.

To expand the analysis and check the robustness of our results, we generate two Erdős-Rényi random directed networks based on Honda's supply network (realistic supply networks exhibit scale-free characteristics). To make random networks comparable to realistic supply networks, we follow Kim et al. (2015) and adopt the NK model. Specifically, we create random networks as the supply base of the focal firm (i.e., Honda). The number of nodes for each supply base (i.e., random network) is 95 without Honda. The number of directed edges is 96, equal to the number of edges

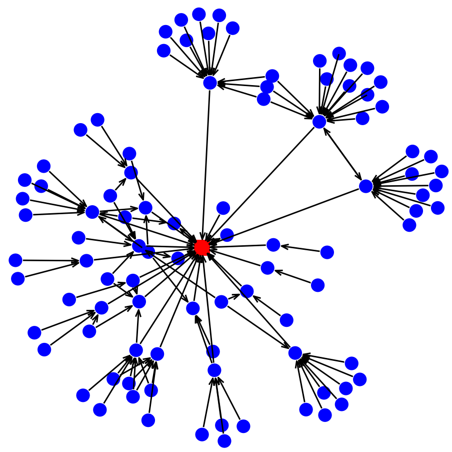


Figure 4.2: Supply networks of Honda (left) and Toyota (right)

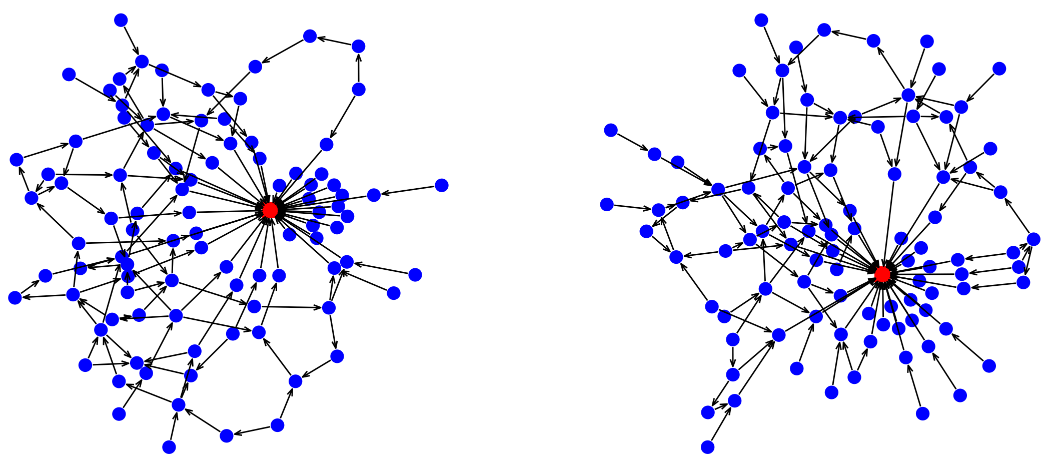


Figure 4.3: Two random networks generated from Honda's network

in Honda's supply network excluding those directed to Honda. We add Honda to the randomly structured supply base and create directed edges from all nodes without outgoing edges to Honda. Figure 4.3 displays the two derived random networks.

4.5.2 Experimental Design and Parameters

The literature review and analytical analysis suggest potential factors that can affect supply network resilience learning. We design two experiments to respectively determine (1) the effect of suppliers' learning rates on network learning and the contingent role of risk, and (2) the effect of a supplier's structural position in the network on network learning. In the first experiment, all suppliers learn at the same rates $a_i = a$ and $b_i = b$. In the second experiment, we vary learning rates across suppliers. In Experiment II, we select some nodes to learn at the faster rates a_x and b_x (larger than a and b, respectively) while the unselected nodes learn at baseline rates a and b. In both experiments, we consider high (75%), medium (50%), and low (25%) probabilities of diffusion and disruption. Since a low (high) β implies rare (frequent) disruption while low (high) γ implies severe (trivial) disruption, we consider rare-severe and frequent-trivial disruptions in both experiments, by setting $\beta = \gamma$. These two kinds of disruptions commonly appear in practice, whereas rare-trivial disruption is of less practical interest and frequent-severe disruption is too destructive to be realistic.

Design of Experiment I – All Suppliers Learn at the Same Rates.

This experimental design investigates how different learning rates and risk levels affect network learning, but all suppliers are homogeneous in terms of learning rates. Experiment I is a full factorial design with five factors (see Table 4.2). In particular, the same supplier learning rates $a_i = a$ and $b_i = b$ are determined by numerically analyzing the supplier learning curves (Equations (4.6) and (4.7)). We pick representative a and b that are neither too large nor too small to avoid quick convergence of β_i^t and γ_i^t, and to tease out the effects of supplier learning on supply network resilience learning as well. The choice has been numerically verified to generate most distinctive β_i^t and γ_i^t along time. The agent-based simulation on each experimental treatment is replicated 50 times, with each replicate lasting for $T = 1000$ time units. The choice of the simulation length of 1000 time units is used because the curve of the PDP becomes stable and goes towards the analytical infinite-time PDP by $t = 1000$ (see visualization in the next Section), without any abnormal pattern or deviation. Em-

Table 4.2: Settings for Experiments I and II

	Experiment I	Experiment II
Networks	All four networks	All four networks
Diffusion Probability α	25%, 50%, 75%	25%, 50%, 75%
Disruption Probability β	25%, 50%, 75%	25%, 50%, 75%
Learning-to-Prevent a	0, 0.2, 0.5, 1	N/A
Learning-to-Recover b	0, 0.01, 0.05, 0.1	N/A
Selected nodes learn at (a_x, b_x)	N/A	$(0.2, 0.01)$, $(0.5, 0.05)$, $(1, 0.1)$
Unselected nodes learn at (a, b)	N/A	$(0, 0)$, $(0.2, 0.01)$, $(0.5, 0.05)$
		$a_x > a$ and $b_x > b$
Node selection criterion	N/A	Highest 25% degree centrality vs. Random other 25%
Simulation length T	1000	1000
Replicate	50	50

pirical examination demonstrates a good statistical power of 50 replicates. Column 1 in Table 4.2 summarizes the parameter settings for Experiment I.

Design of Experiment II – Some Suppliers Learn Faster.

Experiment II allows some suppliers to learn at faster rates than other suppliers. It examines the effect of a supplier's network position on network learning, where more centrally located suppliers can have a different effect than non-central suppliers. The selection criterion includes two types of suppliers – "central nodes" that are in the top 25% of degree centrality, and "non-central nodes" that are other random 25% as the control. The degree centrality of a node is measured as the total number of incoming and outgoing edges through the node. We examine central nodes since their central positions can grant them more importance (Borgatti and Li, 2009) in affecting supply network resilience. Degree centrality directly relates to a supplier's vulnerability (the number of neighbors that can diffuse a disruption to the supplier) and the supplier's ability to disrupt others (the number of neighbors that a supplier can diffuse a disruption to).

The selected suppliers learn at increased rates $a_x > a$ and/or $b_x > b$, while the unselected suppliers learn at baseline rates a and/or b. We adopt the same values used in Experiment I for a, b, a_x and b_x. To highlight the role of a supplier's central position, we reduce factor combinations through pairing b and b_x with a and a_x, respectively. Column 2 in Table 4.2 summarizes the parameter settings for Experiment II.

4.6 Data Analytics and Data-Driven Prescription

The data for analysis comes from the agent-based simulation that uses NetLogo v6.0.4. The simulation generates the supply network's PDP over time. This section visualizes the PDP, presents the data analytics model, conducts an empirical analysis, and proposes data-driven prescriptions.

4.6.1 Visualization of Supply Network Resilience

Data visualization (often the first step in a data-driven exploratory study) provides the initial insights into supply network resilience learning. Figure 4.4 shows the mean PDP curves for Honda's supply network under the parameter settings $(\alpha, \beta) = (25\%, 25\%)$ in Experiment I. The left portion of Figure 4.4 shows all 50 PDP curves (semi-transparent) and the corresponding mean PDP (solid red curve), where all suppliers have the learning rates of $(a, b) = (0.2, 0.05)$. The right portion of Figure 4.4 shows the mean PDP for the different levels of a and b under the same level of risk. In general, the visual analysis shows a supply network resilience learning effect – the mean PDP decreases over time due to suppliers' learning from their disruption experiences.

The visual analysis also shows that PDP curves differ in their downward curvatures, with respect to different combinations of supplier learning rates. Broadly

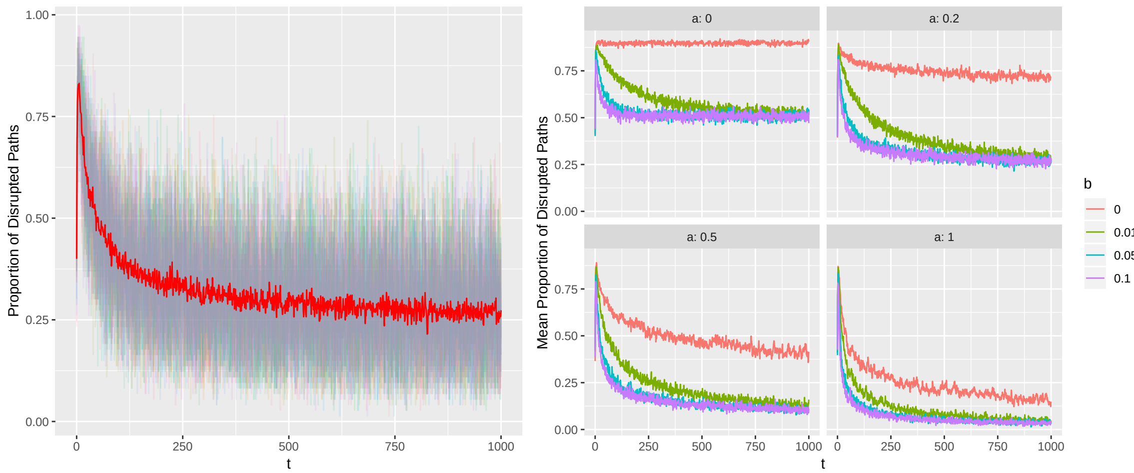


Figure 4.4: Mean PDP curves for Honda's supply network in Experiment I

speaking, this suggests that supplier learning influences supply network resilience learning, and various contingencies influence network learning. Appendix B.2 also provides visual comparison across different risk settings, which further suggests that different network contingencies influence network learning.

Figure 4.5 gives a visual analysis of Experiment II (some suppliers learn faster), and shows the mean PDP for Honda's supply network under different risk levels (α, β) and different learning rates. The red benchmark curve comes from Experiment I where suppliers learn at the same baseline rates (a, b). The computational model shows improved network learning when some (not necessarily central) suppliers learn at higher rates. For instance, blue and green curves are below the red benchmark. We aim to explore which suppliers play a more critical role in supply network resilience learning. Figure 4.5 also suggests that risks play a contingent role, since the blue curve is not always below the green one across different cells. The first few time points constitute the warming-up period, during which risk propagates but suppliers are not learning enough to mitigate risks.

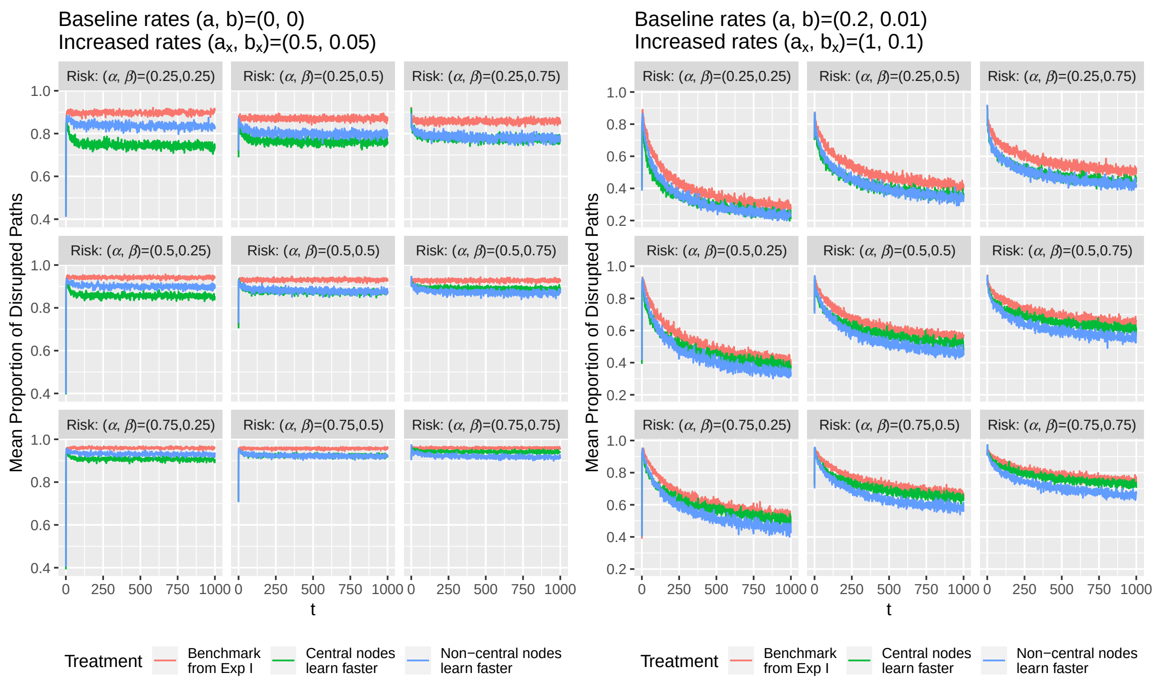


Figure 4.5: Mean PDP curves for Honda's supply network in Experiment II

4.6.2 Data and Empirical Model

The data has a multilevel structure since we investigate the effect of supplier learning at the node level on supply network resilience learning at the network level. Hence, multilevel models naturally fit into the analysis. Multilevel models, also known as hierarchical models or the mixed-effects models (Garson, 2012), are appropriate for research designs with data at more multiple levels (i.e., nested data) (Tabachnick et al., 2007). In our context, suppliers are nested within the (four) supply networks. Therefore, we adopt hierarchical models with two levels (node- and network-level).

Due to a lack of theory that accurately characterizes the curve of the supply network resilience learning (i.e., the PDP curve), we compute the area under the PDP curve (a more nonparametric approach). We follow Zobel and Khansa (2012) and Zobel (2014)'s approach to compute the area under curve ("AUC") as a proxy of the supply network resilience. A lower AUC indicates a more resilient supply network and better network learning. The AUC is computed as

$$\mathrm{AUC}_i^T = \sum_{t=1}^{T} \mathrm{PDP}_i^t - (\mathrm{PDP}_i^1 + \mathrm{PDP}_i^T)/2 \tag{4.18}$$

where i stands for one of the networks, T is the simulation length for AUC, and PDP_i^t is the PDP of network i at time t. In the main analysis, we pick $T = 1000$. Noticing that the temporal length affects the curve shape and AUC, we examine $T = 500$ as a robustness check (see Appendix B.3).

The network-level model is an intercept-only model in the form of

$$\mathrm{AUC}_{ij}^{1000} = \delta_{0i} + \varepsilon_{ij} \tag{4.19}$$

where i stands for one of the four networks, j stands for the individual simulation replicates, δ_{0i} is the mean of AUC for the network i, and ε_{ij} is the random error.

Due to the exploratory nature of the analysis, we assume a parsimonious linear relationship between supplier characteristics and the intercept (i.e., individual network's mean of the AUC). For Experiment I, the following node-level model captures the cross-level relationships.

$$\begin{aligned}\delta_{0i} = \eta_0 &+ \eta_1\alpha + \eta_2\beta + \eta_3 a + \eta_4 b + \eta_5\alpha a + \eta_6\alpha b + \eta_7\beta a + \eta_8\beta b + \eta_9\alpha\beta + \eta_{10}ab \\ &+ \eta_{11}\alpha\beta a + \eta_{12}\alpha\beta b + \eta_{13}\alpha ab + \eta_{14}\beta ab + \eta_{15}\alpha\beta ab + b_{0i}\end{aligned} \tag{4.20}$$

where b_{0i} is the random effect of δ_{0i} from four different networks and η's are the estimated parameters. Notice from the analytical analysis and visualization that the PDP behavior differs between $a = 0$ and $a \neq 0$. To avoid data distortion, we filter out $a = 0$ or $b = 0$ to improve the validity of model fitting and meanwhile to maintain a balanced design.

For Experiment II, due to paired (a, b) and (a_x, b_x), we show only a and a_x in the expressions to avoid the fixed-effect model matrix to be rank deficient (entering b and b_x instead of a and a_x returns similar results). As the number of factors increases in the model, we correspondingly express the full model at the node level in a following concise way.

$$\delta_{0i} = \boldsymbol{\eta}[(\alpha + 1) \times (\beta + 1) \times (\text{Central} + 1) \times (a + 1) \times (a_x + 1)] + b_{0i} \tag{4.21}$$

where Central is a binary categorical variable (reference: non-central nodes, i.e., if central nodes are selected to learn faster, Central $= 1$), $\boldsymbol{\eta}$ is a vector of the parameters, and b_{0i} is the random effect. In general, b_{0i} captures supply network randomness like the network topology. The fixed effects due to the risk on the suppliers and supplier characteristics are reflected as the η-coefficients.

4.6.3 Analysis, Results, and Data-Driven Prescriptions

Following Aguinis et al. (2013), we integrate node- and network-level models and fit the composite model to the data to examine the contingencies under which supply network resilience learning can improve more. To conform to the normality assumption of the multilevel models, we log-transform the dependent variable AUC. Diagnostics indicate that model assumptions are satisfied.

Analysis of Experiment I.

The purpose of Experiment I is to identify the contingencies under which suppliers' learning-to-prevent and/or learning-to-recover improve supply network resilience learning. Table 4.3 presents the analysis results for Experiment I where all suppliers learn at the same rate. Models (1) through (3) are mixed-effects models (i.e., multilevel models) using all 4 networks, while we additionally regress ln(AUC) against the predictors for each individual network as the robustness check (models (4) through

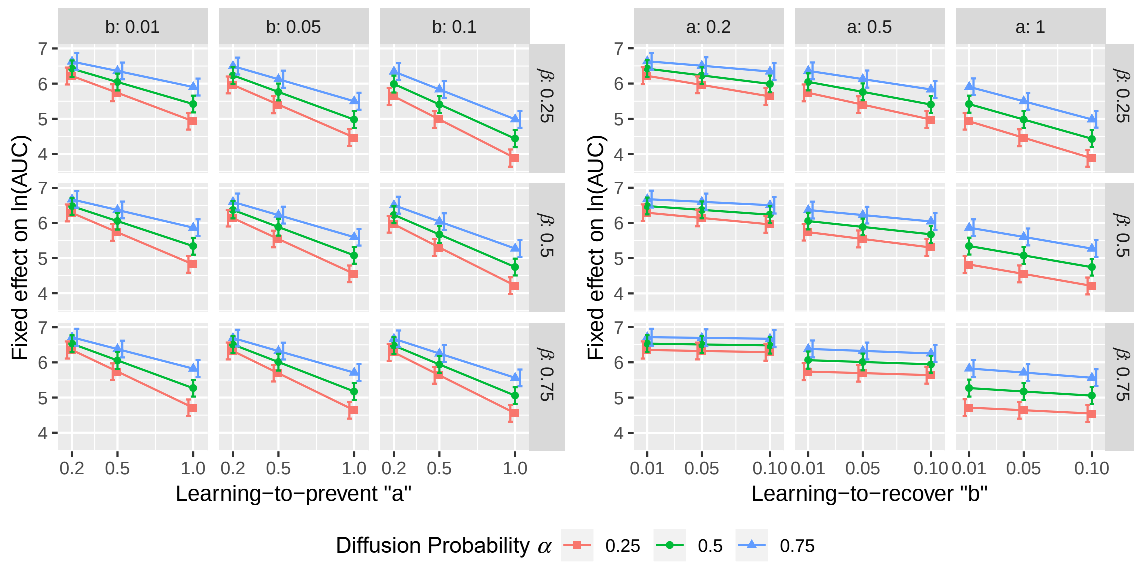


Figure 4.6: Interaction effects between risk and supplier learning in Experiment I

(7), OLS diagnostics checked).

Interaction terms are entered hierarchically. Model selection indicates that model (3) (the full model) has the most explanatory power (highest log-likelihood and lowest AIC and BIC). In general, all 7 models in Table 4.3 give consistent estimates in terms of direction, magnitude, and significance.

A lower ln(AUC) is desirable and indicates a lower PDP peak and/or a steeper downward PDP curve, hence an improved supply network resilience learning. Both learning-to-prevent and learning-to-recover, as suggested by the main effects, significantly reduce AUC. In the following, we focus on the interaction terms displayed as Figure 4.6 with 90% confidence intervals obtain through $10,000$ times of bootstrapping.

First, we focus on the interactions between risk and supplier learning modes. The left panel of Figure 4.6 shows that suppliers' learning-to-prevent ("a") improves network learning more under a lower diffusion probability α, because the slope of the red line is steeper than green and blue ones in every cell with non-overlapping error bars. The effect of learning-to-prevent is non-distinguishable across different disruption probabilities β. In the right panel, the suppliers' learning-to-recover ("b") improves network learning more under a lower disruption probability β because the slopes for the smaller β are steeper with non-overlapping error bars, but there is no significant effect across different diffusion rates α. Collectively these results suggest the following propositions.

Proposition 6 Suppliers' learning-to-prevent improves supply network resilience learning more under a lower diffusion probability.

Proposition 7 Suppliers' learning-to-recover improves supply network resilience learning more under a lower disruption probability.

Second, we examine the interaction between two modes of supplier learning. The suppliers' learning-to-prevent and learning-to-recover exhibit a synergistic effect because improving both modes reduces AUC quicker (-7.675^{***}). The synergistic effect is more prominent under a higher diffusion (-5.358^{***}) but a lower disruption probability (9.617^{***}). Figure 4.6 also illustrates this synergistic effect and the associated contingency. In Figure 4.6, we can compare the same colored lines across columns in each row. When the focal firm and suppliers are capable to learn to both prevent and recover, they can better take advantage of the synergistic effect.

Proposition 8 Suppliers with both high learning to prevent and recover have a positive synergistic effect on increasing supply network resilience learning, especially under a high diffusion probability and a low disruption probability.

Table 4.3: Experiment I (same supplier learning rates) results

	Mixed-Effects			*OLS* (robustness check)			
DV: ln(AUC)	(1) All	(2) All	(3) All	(4) Honda	(5) Toyota	(6) Rnd.1	(7) Rnd.2
Constant	6.055***	6.581***	6.373***	6.398***	6.456***	6.329***	6.309***
	(0.039)	(0.039)	(0.042)	(0.025)	(0.024)	(0.034)	(0.034)
Diffusion (α)	1.220***	0.468***	0.533***	0.476***	0.471***	0.606***	0.578***
	(0.007)	(0.011)	(0.030)	(0.046)	(0.045)	(0.063)	(0.063)
Disruption (β)	0.400***	0.099***	0.419***	0.390***	0.370***	0.460***	0.455***
	(0.007)	(0.011)	(0.030)	(0.046)	(0.045)	(0.063)	(0.063)
Prevent (a)	−1.568***	−2.171***	−1.609***	−1.446***	−1.325***	−1.826***	−1.837***
	(0.004)	(0.011)	(0.029)	(0.044)	(0.043)	(0.060)	(0.060)
Recover (b)	−3.833***	−10.591***	−10.204***	−8.797***	−8.127***	−12.056***	−11.837***
	(0.038)	(0.109)	(0.289)	(0.444)	(0.434)	(0.606)	(0.609)
$\alpha \times a$		1.525***	1.228***	0.747***	0.758***	1.675***	1.733***
		(0.015)	(0.053)	(0.081)	(0.079)	(0.111)	(0.111)
$\alpha \times b$		2.610***	10.637***	7.717***	7.886***	13.835***	13.109***
		(0.148)	(0.534)	(0.823)	(0.803)	(1.123)	(1.127)
$\beta \times a$		−0.320***	−1.216***	−1.067***	−1.118***	−1.334***	−1.343***
		(0.015)	(0.053)	(0.081)	(0.079)	(0.111)	(0.111)
$\beta \times b$		10.907***	12.569***	10.841***	10.057***	14.936***	14.444***
		(0.148)	(0.534)	(0.823)	(0.803)	(1.123)	(1.127)
$\alpha \times \beta$			−0.268***	−0.233**	−0.266**	−0.300*	−0.273*
			(0.056)	(0.086)	(0.083)	(0.117)	(0.117)
$a \times b$			−7.675***	−9.086***	−10.611***	−5.384***	−5.620***
			(0.508)	(0.783)	(0.764)	(1.068)	(1.072)
$\alpha \times \beta \times a$			0.920***	1.020***	1.162***	0.752***	0.744***
			(0.098)	(0.151)	(0.147)	(0.205)	(0.206)
$\alpha \times \beta \times b$			−12.592***	−9.077***	−9.379***	−16.651***	−15.261***
			(0.990)	(1.524)	(1.488)	(2.079)	(2.087)
$\alpha \times a \times b$			−5.358***	2.544†	4.029**	−14.309***	−13.694***
			(0.941)	(1.449)	(1.415)	(1.977)	(1.985)
$\beta \times a \times b$			9.617***	9.975***	11.629***	7.981***	8.882***
			(0.941)	(1.449)	(1.415)	(1.977)	(1.985)
$\alpha \times \beta \times a \times b$			2.571	−3.937	−5.419*	11.007**	8.632*
			(1.743)	(2.683)	(2.619)	(3.660)	(3.676)
Observations	28,800	28,800	28,800	7,200	7,200	7,200	7,200
Adjusted R^2				0.954	0.951	0.939	0.938
Log Likelihood	−1,115.3	5,213.0	8,074.8				
AIC	2,244.6	−10,404	−16,114				
BIC	2,302.5	−10,313	−15,965				
Resi. SE (df = 7184)				0.141	0.137	0.192	0.193
F Stat (df = 15; 7184)				9,978.9***	9,219.2***	7,369.3***	7,280.6***

Note: †p < .1; *p < .05; **p < .01; ***p < .001
Standard errors of coefficients are displayed in the parentheses.

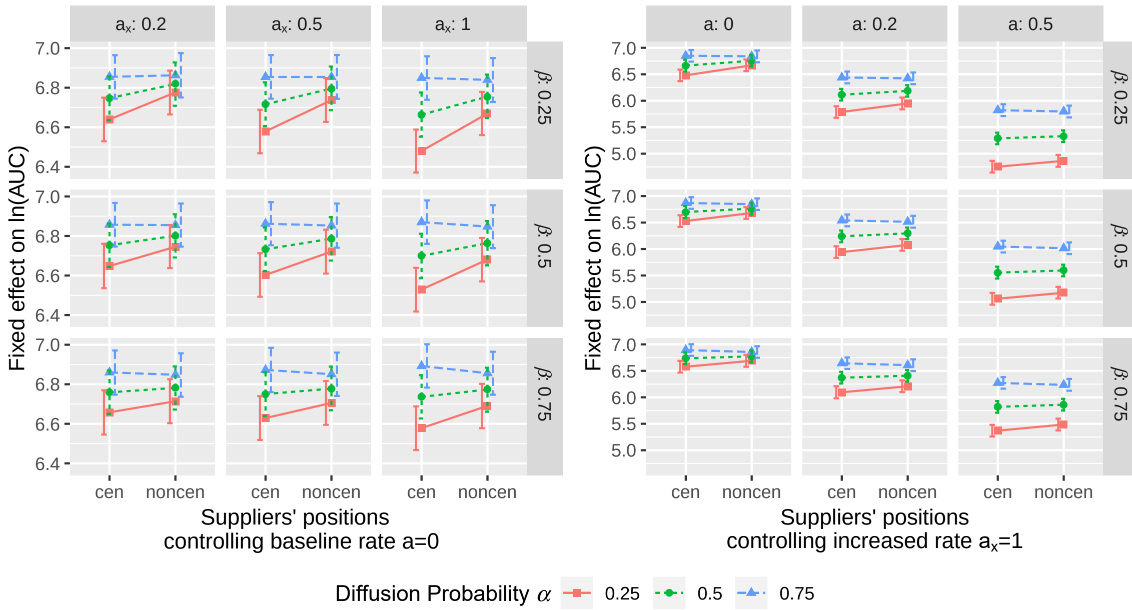


Figure 4.7: Interaction effects between centrality, risk, and supplier learning in Experiment II

Analysis of Experiment II.

Experiment II examines the effect of the suppliers' network centrality on supply network resilience learning. We purposefully pair b and b_x to a and a_x respectively to reduce factor combinations and highlight the importance of a supplier's structural position. Table 4.4 and Figure 4.7 present the results.

Models (8) through (12) give the mixed-effects models. Models (11) and (12) analyze the special situations of $a = 0$ and $a_x = 1$, respectively, as the robustness check. Models (13) and (14) are OLS regression for further robustness check (to save space, only Honda's supply network and random network 1 are included). Interaction terms are entered hierarchically. Model selection favors the full model (10) with the highest log-likelihood and lowest AIC and BIC. In general, for highly significant terms, the estimates are consistent in the direction across models. Due to the involvement of high-order interaction terms, we focus on Figure 4.7 with the 90% bootstrapped confidence intervals to interpret the results.

First, we examine the interaction between risk and suppliers' structural positions. Both diffusion probability α and disruption probability β moderate the importance of suppliers' positions in supply network resilience learning. Central suppliers are more important under a lower diffusion rate (slopes are more positive for a lower α in each cell; an estimate of 0.295^{***} in model (10)). Similarly, central nodes are more important under a lower disruption probability (same-colored slopes are more positive for a lower β given a column; an estimate of 0.233^{***} in model (10)). Moreover, lower diffusion and lower disruption rates increase the importance of central nodes (-0.265^{***}).

Proposition 9 Central suppliers' learning increases supply network resilience learning more under a low disruption probability or a low diffusion probability.

Proposition 9 implies that non-central suppliers may be more important under higher diffusion and disruption probabilities. Next, we investigate the effect of the increased learning rates (i.e., a_x relative to a) on central and non-central nodes. From Figure 4.7, as the gap between a and a_x increases, positive (negative) slopes become more positive (negative). Hence, we conclude that non-central nodes are more important when the rate increment is higher and under higher diffusion and disruption rates (0.158^{**} and $-0.693^{\dagger}$ in Model (10)).

Proposition 10 Non-central suppliers' learning increases supply network resilience learning more when both disruption and diffusion probabilities are high, especially under a higher increase in supplier learning rates from (a, b) to (a_x, b_x).

Table 4.4: Experiment II (some suppliers learn faster) results

	Mixed-Effects					*OLS* (robustness check)	
DV: ln(AUC)	(8) All	(9) All	(10) All	(11) All	(12) All	(13) Honda	(14) Rnd.1
Constant	6.183***	6.657***	6.829***	6.753***	6.563***	6.805***	6.856***
	(0.030)	(0.030)	(0.033)	(0.012)	(0.043)	(0.016)	(0.020)
Diffusion (α)	0.868***	0.278***	0.066*	0.151***	0.357***	0.094**	0.028
	(0.005)	(0.010)	(0.028)	(0.011)	(0.026)	(0.029)	(0.036)
Disruption (β)	0.385***	0.021*	−0.229***	−0.125***	0.063*	−0.222***	−0.237***
	(0.005)	(0.010)	(0.028)	(0.011)	(0.026)	(0.029)	(0.036)
Central	−0.051***	−0.178***	−0.241***	−0.247***	−0.340***	−0.196***	−0.265***
	(0.002)	(0.010)	(0.021)	(0.009)	(0.020)	(0.022)	(0.028)
Baseline rate (a)	−2.321***	−4.888***	−8.978***		−5.037***	−8.063***	−10.171***
	(0.007)	(0.019)	(0.153)		(0.045)	(0.158)	(0.199)
Increased rate (a_x)	−0.051***	−0.052***	−0.255***	−0.013		−0.231***	−0.295***
	(0.004)	(0.011)	(0.022)	(0.009)		(0.022)	(0.028)
$\alpha \times a$		2.839***	8.183***		3.383***	6.290***	10.215***
		(0.026)	(0.284)		(0.083)	(0.292)	(0.368)
$\alpha \times a_x$		0.037*	0.279***	0.007		0.200***	0.373***
		(0.015)	(0.040)	(0.017)		(0.041)	(0.052)
$\beta \times a$		2.249***	8.393***		2.715***	7.552***	9.449***
		(0.026)	(0.284)		(0.083)	(0.292)	(0.368)
$\beta \times a_x$		−0.015	0.273***	−0.058***		0.211***	0.361***
		(0.015)	(0.040)	(0.017)		(0.041)	(0.052)
$\alpha \times$ Centr		0.172***	0.295***	0.307***	0.448***	0.192***	0.368***
		(0.014)	(0.039)	(0.016)	(0.037)	(0.041)	(0.051)
$\beta \times$ Centr		0.097***	0.233***	0.232***	0.212***	0.230***	0.219***
		(0.014)	(0.039)	(0.016)	(0.037)	(0.041)	(0.051)
Centr $\times a$		0.198***	0.319		0.337***	0.771***	−0.188
		(0.027)	(0.217)		(0.063)	(0.223)	(0.281)
Centr $\times a_x$		−0.113***	−0.104***	−0.085***		0.003	−0.204***
		(0.016)	(0.031)	(0.013)		(0.032)	(0.040)
$\alpha \times \beta$			0.248***	0.150***	−0.034	0.246***	0.254***
			(0.052)	(0.021)	(0.048)	(0.053)	(0.067)
$a \times a_x$			3.912***			3.297***	4.731***
			(0.161)			(0.166)	(0.209)
$\alpha \times$ Centr $\times a$		−0.099**	−0.300		−0.391***	−0.163	−0.353
		(0.037)	(0.401)		(0.117)	(0.413)	(0.520)
$\alpha \times$ Centr $\times a_x$		0.173***	0.158**	0.119***		0.047	0.260***
		(0.021)	(0.057)	(0.025)		(0.058)	(0.074)
$\beta \times$ Centr $\times a$		−0.201***	−0.693†		−0.481***	−1.022*	−0.378
		(0.037)	(0.401)		(0.117)	(0.413)	(0.520)
$\beta \times$ Centr $\times a_x$		0.011	−0.021	−0.017		0.026	−0.069
		(0.021)	(0.057)	(0.025)		(0.058)	(0.074)
$\alpha \times \beta \times$ Centr			−0.265***	−0.266***	−0.213**	−0.203**	−0.300**
			(0.073)	(0.030)	(0.068)	(0.075)	(0.095)
Centr $\times a \times a_x$			0.029			−0.399†	0.501†
			(0.228)			(0.234)	(0.295)
...... Other higher-order terms omitted to save space. They are insignificant or not related to Central							
Observations	21,600	21,600	21,600	10,800	10,800	5,400	5,400
Adjusted R^2						0.992	0.991
Log Likelihood	8,722.7	22,113	23,624	21,665	9,889.0		
Akaike Inf. Crit.	−17,429	−44,185	−47,181	−43,294	−19,742		
Bayesian Inf. Crit.	−17,366	−44,026	−46,909	−43,163	−19,611		
F Stat (df = 31; 5368)						20,480***	18,722***

Note: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$
S.E. of coefficients displayed in parentheses.

Robustness Check and *Post Hoc* Analysis.

The main analysis includes several robustness checks, such as random networks, OLS regression, model selection, and learning rates ($a = 0$ or $a_x = 1$). Appendix B.3 gives an additional robustness check for the simulation length of 500. The tables and figures are constructed in the same manner as the main analysis. The results are consistent with the main analysis, suggesting that the length of the finite time horizon that we consider does not affect the results.

Appendix B.4 gives an additional *post hoc* analysis, which further investigates the role of central suppliers. The *post hoc* analysis examines the difference between in-degree and out-degree centralities of the suppliers. In-degree centrality measures the number of a supplier's incoming edges and out-degree measures the number of a supplier's outgoing edges. In-degree reflects the vulnerability of a supplier. In other words, a supplier with a high in-degree is more susceptible to risk propagation from its suppliers. On the contrary, out-degree reflects the potential of a supplier to transmit a disruption to others. A supplier with a high out-degree can more easily diffuse risks to its buyers. Hence, the *post hoc* analysis examines the differences between reducing risks at the suppliers that diffuse a disruption versus fortifying susceptible suppliers against risks.

The experiment for the *post hoc* analysis is similar to Experiment II, with the node selection criterion (for faster learning) now being the nodes at top 25% out-degree vs. in-degree. We build a similar mixed-effects model in the following composite form:

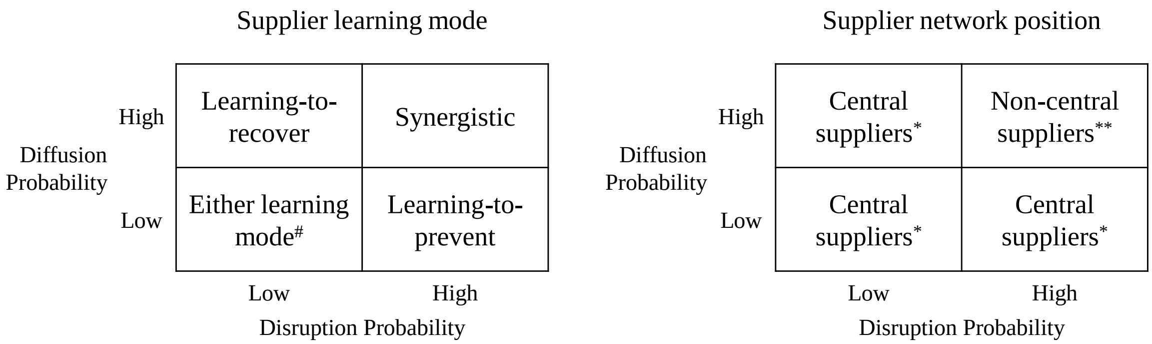
$$\mathrm{AUC}_{ij}^{1000} = \boldsymbol{\eta}[(\alpha+1)\times(\beta+1)\times(\mathrm{Out}+1)\times(a+a_x+1)] + b_{0i} + \varepsilon_{ij} \tag{4.22}$$

where Out is a binary categorical variable (Out $= 1$ if nodes with higher out-degree centralities learn faster). Table B.3 and Figure B.3 in Appendix B.4 indicate that increasing learning rates for suppliers of higher out-degree always improves supply network resilience learning more than in-degree, regardless of the other treatment levels. In other words, it is always better to try to eliminate risks from their sources rather than to help vulnerable nodes learn faster.

4.7 Implications, Discussion, and Conclusion

Increasingly, a (focal) firm's performance depends on the performance of its supply network. The recent example of Ford's production of the F-150 trucks getting disrupted by its supplier illustrates how firms increasingly become susceptible to disruptions from their supply bases. Supply managers face a daunting task of managing operational risks in their supply networks. Companies like Ford want to improve the resilience in their supply networks and reduce risks, so that their operations do not get disrupted. This research examines the question of *how the supply network learns from suppliers' disruptions*. The analysis takes a multilevel approach that bridges node-level and network-level learning, and characterizes the contingencies that influence supply network resilience learning. The analysis shows that three factors influence resilience at the network level: the suppliers' learning rates, the operational risk, and the suppliers' structural positions in the network. These factors interact to influence supply network resilience learning.

A supply network learns through its suppliers to reduce disruption and improve resilience. Network learning occurs with a reduction of the proportion of disrupted paths (PDP) in the supply network. In the long run (infinite time), the suppliers' learning-to-prevent will result in a disruption-free supply network. Focusing on learning-to-prevent may overall be an effective strategy for a firm that operates in a slow clock-speed industry where its supply base does not change much. However, firms and their supplier relationships increasingly change over time, so performance over a finite time horizon becomes more relevant. In this situation, the suppliers'



#The synergistic effect in this situation can also improve supply network resilience learning more
*Better to increase learning rates for suppliers with higher out-degree centralities (*post hoc* results)
**Especially under a higher increase in the supplier learning rates from (a, b) to (a_x, b_x)

Figure 4.8: Data-driven strategies that improve supply network resilience learning

learning-to-prevent improves network learning more under a lower diffusion probability, whereas learning-to-recover improves network learning more with a lower probability of a disruption. The two learning modes can also exert a synergistic effect on network learning, especially when the suppliers face a high probability of diffusion. Furthermore, centrally located suppliers' learning rates influence network learning more, especially under a lower disruption or diffusion rate. However, non-central nodes have a bigger impact on supply network resilience learning when both the diffusion and disruption rates are high. Notice that improving supplier learning under all conditions improves supply network resilience learning. However, by showing the contingencies that moderate network-level learning, this study suggests strategies to *more effectively* enhance network learning. Since firms have limited resources, they need to know where to allocate these scarce resources to improve network resilience. Figure 4.8 summarizes the findings from this study and suggests strategic guidelines to improve supply network resilience learning.

The analytical model and data-driven findings have many practical implications for managing operational risks in the supply networks. For both existing and new incoming suppliers, the supply manager can assess the risk level of diffusion and disruption, along with the suppliers' positions in the network, to develop plans accordingly to improve network resilience. In the risk quadrant (Figure 4.8), first, if suppliers introduce high-diffusion and high-disruption risks into the supply network, supply managers need to help the suppliers learn how to both prevent and recover from the risk – our results suggest that focusing on only one mode of learning for these suppliers may not lead to the best results for the overall network. In addition, although central suppliers play a more critical role in supply network resilience, managers should not overlook non-central suppliers, especially under the high disruption and diffusion risk. That is, supply managers need to focus on suppliers that face a high risk of disruption and diffusion even if they are not centrally located in the supply network. Second, supply managers should more help central suppliers that face a high diffusion risk but low disruption risk focus on learning to recover from a disruption. In contrast, for suppliers with a low diffusion risk and high disruption risk, they need to focus on learning to prevent a disruption. For the last cell of the risk quadrant, supply managers should be cautious to quickly discount suppliers that have low diffusion and disruption rates. The analysis shows that for these suppliers, risk propagation still leads to more than 80% of disruptions (see visualization of PDP curves under $(\alpha, \beta) = (0.25, 0.25)$ in Appendix B.2) in the network. Low risk levels should never be overlooked and either mode of supplier learning (or the synergistic learning) can improve network resilience.

Finally, this research suggests the importance of taking a multilevel perspective to manage the supply network. We believe that this general approach can be applied to other areas in the supply network management. For instance, quality management has a strong connection to the organizational learning field. As a supply manager looks to improve quality performance from the supply network, he/she can potentially apply the method and results from this study. More generally, we believe this study helps to understand how supply managers can better manage their supply networks.

This study has several limitations, which could affect the conclusions. First, it does not consider organizational forgetting and that suppliers might not retain what they have learned. In addition, suppliers can potentially learn from their neighbors and learn at different rates, forming more complex learning curves than given in Equations (4.6) and (4.7). Moreover, the network-level resilience metric may be extended to path flow amount rather than path status (i.e., healthy or not), by incorporating quantities of actual physical material flows. Other different metrics (e.g., time-to-recover) and other model formation (e.g., continuous-time model) may be adopted as well.

Furthermore, due to a lack of strong theory, this study assumes supply network resilience learning is related to supplier learning, builds up models of the complex system, and relies on the agent-based simulation to investigate the network-level learning. Although we base the model on actual automobile supply networks, secondary data analysis on operational risks and network learning can further verify the model and propositions.

Finally, it is critical to understand the "knowledge" related to supply network resilience learning, to further strengthen the concept. Learning involves the "processes of creating, retaining and transferring knowledge" (Argote and Hora, 2017, p.579). This study focuses on the observed outcome of network learning (the PDP curves) without clarifying the behavior of knowledge within and/or across supply networks. We leave the discussion of knowledge to future research, and call for more aspects towards network learning to deepen our understanding of this previously neglected area.

Chapter 5

Concluding Remarks

5.1 Contributions and Implications

Most of the supply chain and operations literature focuses on how the individual firm can improve their performance. This dissertation takes a broader network perspective and proposes that operational performance of a firm also depends on the network that it resides in or its network structural position. There are many emerging areas in the field of supply chain management for researchers to explore from a network perspective. This dissertation examines three areas of supply chain operational performance from a network perspective.

To the best of our knowledge, Essay 1 is the first to understand the effects of structural embeddedness on warehouse performance in the context of a warehouse network. Few studies, if any, have examined the role of direct vs. indirect ties on warehouse inventory efficiency. Understanding the role and performance of a warehouse in the network will become increasingly important as firms rely more on their warehouse and distribution networks as a source of competitive advantage. In addition to using traditional benchmarking strategies in assessing a warehouse's performance, managers should assess the warehouse relative to its position within its network. Failure to do so could lead to biased assessments and inaccurate estimation.

Essay 2 contributes to the field of supply network resilience by characterizing the optimal resilience investment strategies. Drawing on the extant research on network flow optimization, Essay 2 provides strategies for investing limited resources in supply chain resilience. The results give insights into making investments in critical nodes versus critical paths. Essay 2 also generates and generalizes data-driven prescriptions that fill in the void of resilience investment in the supply chain network. This network perspective also examines contingencies related to different disruption frequencies and routing mechanisms.

Essay 3 fills in a large theoretical void regarding the relationship between network-level resilience learning and node-level disruption risks. Essay 3 makes a multilevel connection between supplier learning from disruptions and network learning through an agent-based model and simulation. In addition, the contingent role of the risk is highlighted. These contingencies suggest propositions that offer strategic guidelines on how to more effectively improve supply network resilience. Due to a lack of theory and the analytical difficulty in characterizing network-level learning, Essay 3 contributes to taking a data-driven approach to understand operational risks in the supply network.

This dissertation also generates managerial implications. In a broader sense, results of the three essays are applicable to different network structures or situations. Essay 1 suggests that managers should assess a warehouse's inventory efficiency relative to its position in the network, as a complement to the traditional benchmarking strategy. Essay 2 proposes investment guidelines based on the generic model. Essay 3 suggests optimal learning modes for suppliers and guides the focal firm to facilitate learning for key suppliers. Generally speaking, managers can benefit from the suggestions to make better operational decisions.

This dissertation is motivated by and grounded in real supply chains. Essays 1 and 2 use the network data and the operational context related to a world leading logistics management company, while Essay 3 analyzes Honda's and Toyota's supply networks. The research questions are motivated by the practical difficulties that the companies face. Hence, results and implications are meaningful and useful to the companies. Should opportunities arise to implement our suggestions in the logistics management company's actual operations, the validity and robustness of our results can be further demonstrated.

5.2 Future Research Direction

This dissertation has a few limitations that suggest new research directions to explore further. First, the issue of causality long exists in the supply chain and operations management field and has sparked vigorous debate (Ho et al., 2017). The analysis in Essay 1 is time-series cross-sectional and is restricted by the available dataset (one warehouse network). It is therefore hard to derive the strong causality between structural embeddedness and warehouse inventory efficiency. Future research can explore effective ways to address the causality of the network structural effects.

Second, analytical and simulation models tend to make assumptions because of tractability. While the assumptions are grounded in practice, there can still be many extensions. For example, in Essay 3, an important assumption is that suppliers learn to prevent and to recover following the power and exponential curves, respectively. The organizational learning literature suggests that such learning can be much more complicated. Hence, simple forms of learning curve may not comprehensively capture the entire picture. In addition organizations can also "forget" what they have learned. In this sense, future empirical research is strongly called for to verify the propositions from essays in this dissertation. Other extensions like correlated disruptions among suppliers, differentiated learning rates, etc., can all make the model more realistic as well as less tractable.

Finally, the contrast between "node" and "network" is worth the discussion. Specifically, in Essay 1, individual warehouse's performance is examined, while in Essays 2 and 3, the network output and network performance are investigated. Such distinction lies in the contexts, where warehouses in the network are more "autonomous" (each warehouse can influence the transshipment) while the supply network and the distribution network are more "governed" by a focal firm. For instance, in Essay 2, the sourcing and fulfillment in the supply chain network can be determined by the contracts. We shall pay special attention to the context, and determine accordingly whether a specific network framework (e.g., social network analysis) can be applied to the analysis of network operation.

Bibliography

Ackerman, K. B. (1990). Warehouse performance audits. In *Practical Handbook of Warehousing*, pages 113–120. Springer.

Agrawal, A. and Muthulingam, S. (2015). Does organizational forgetting affect vendor quality performance? an empirical investigation. *Manufacturing & Service Operations Management*, 17(3):350–367.

Aguinis, H., Gottfredson, R. K., and Culpepper, S. A. (2013). Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management*, 39(6):1490–1528.

Albert, R., Jeong, H., and Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, 406(6794):378.

Ang, E., Iancu, D. A., and Swinney, R. (2016). Disruption risk and optimal sourcing in multitier supply networks. *Management Science*, 63(8):2397–2419.

Arellano, M. (1987). Practitioners'corner: Computing robust standard errors for within-groups estimators. *Oxford bulletin of Economics and Statistics*, 49(4):431–434.

Arellano, M. (2003). *Panel data econometrics*. Oxford university press.

Argote, L. (2012). *Organizational learning: Creating, retaining and transferring knowledge*. Springer Science & Business Media.

Argote, L. and Epple, D. (1990). Learning curves in manufacturing. *Science*, 247(4945):920–924.

Argote, L. and Hora, M. (2017). Organizational learning and management of technology. *Production and Operations Management*, 26(4):579–590.

Bai, W. and Bilmes, J. (2018). Greed is still good: Maximizing monotone submodular+supermodular (bp) functions. In *International Conference on Machine Learning (ICML)*, Stockholm, Sweden.

Baker, P., Croucher, P., and Rushton, A. (2017). *The Handbook of Logistics and Distribution Management: Understanding the Supply Chain*. Kogan Page Publishers.

Bakshi, N. and Mohan, S. (2015). Mitigating disruption cascades in supply networks. *Preprint at http://faculty. london. edu/nbakshi/Nw. pdf*.

Ball, M. O. (1986). Computational complexity of network reliability analysis: An overview. *IEEE Transactions on Reliability*, 35(3):230–239.

Basole, R. C. and Bellamy, M. A. (2014). Supply network structure, visibility, and risk diffusion: A computational approach. *Decision Sciences*, 45(4):753–789.

Baum, J. A., Cowan, R., and Jonard, N. (2010). Network-independent partner selection and the evolution of innovation networks. *Management science*, 56(11):2094–2110.

Baum, J. A. and Dahlin, K. B. (2007). Aspiration performance and railroads' patterns of learning from train wrecks and crashes. *Organization Science*, 18(3):368–385.

Baumol, W. J. and Ide, E. A. (1956). Variety in retailing. *Management Science*, 3(1):93–101.

Beamon, B. M. (1998). Supply chain design and analysis: Models and methods. *International journal of production economics*, 55(3):281–294.

Bellamy, M. A., Ghosh, S., and Hora, M. (2014). The influence of supply network structure on firm innovation. *Journal of Operations Management*, 32(6):357–373.

Berry, W. L. and Cooper, M. C. (1999). Manufacturing flexibility: methods for measuring the impact of product variety on performance in process industries. *Journal of Operations Management*, 17(2):163–178.

Bimpikis, K., Candogan, O., and Ehsani, S. (2017). Supply disruptions and optimal network structures.

Bimpikis, K., Fearing, D., and Tahbaz-Salehi, A. (2018). Multisourcing and miscoordination in supply chain networks. *Operations Research*, 66(4):1023–1039.

Birge, J. R. and Louveaux, F. (2011). *Introduction to stochastic programming*. Springer Science & Business Media.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences*, 99(suppl 3):7280–7287.

Borgatti, S. P. and Foster, P. C. (2003). The network paradigm in organizational research: A review and typology. *Journal of management*, 29(6):991–1013.

Borgatti, S. P. and Li, X. (2009). On social network analysis in a supply chain context. *Journal of Supply Chain Management*, 45(2):5–22.

Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G. (2009). Network analysis in the social sciences. *science*, 323(5916):892–895.

Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, 17(31):334–355.

Breusch, T. S. and Pagan, A. R. (1980). The lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, 47(1):239–253.

Brinkley, C. (2018). The small world of the alternative food network. *Sustainability*, 10(8):2921.

Cachon, G. P. (1999). Managing supply chain demand variability with scheduled ordering policies. *Management science*, 45(6):843–856.

Carey, M. and Hendrickson, C. (1984). Bounds on expected performance of networks with links subject to failure. *Networks*, 14(3):439–456.

Carpenter, M. A., Li, M., and Jiang, H. (2012). Social network research in organizational contexts: A systematic review of methodological issues and choices. *Journal of Management*, 38(4):1328–1361.

Carter, C. R., Rogers, D. S., and Choi, T. Y. (2015). Toward the theory of the supply chain. *Journal of Supply Chain Management*, 51(2):89–97.

Chandrasekaran, A., Linderman, K., and Sting, F. J. (2018). Avoiding epistemological silos and empirical elephants in om: How to combine empirical and simulation methods? *Journal of Operations Management*, 63:1–5.

Chatfield, D. C., Hayya, J. C., and Cook, D. P. (2013). Stockout propagation and amplification in supply chain inventory systems. *International Journal of Production Research*, 51(5):1491–1507.

Choi, T.-M., Cheng, T., and Zhao, X. (2016). Multi-methodological research in operations management. *Production and Operations Management*, 25(3):379–389.

Choi, T. Y., Dooley, K. J., and Rungtusanatham, M. (2001). Supply networks and complex adaptive systems: control versus emergence. *Journal of operations management*, 19(3):351–366.

Choi, T. Y. and Kim, Y. (2008). Structural embeddedness and supplier management: a network perspective. *Journal of Supply Chain Management*, 44(4):5–13.

Choi, T. Y. and 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. and Wu, Z. (2009a). Taking the leap from dyads to triads: Buyer–supplier relationships in supply networks. *Journal of Purchasing and Supply Management*, 15(4):263–266.

Choi, T. Y. and Wu, Z. (2009b). Triads in supply networks: Theorizing buyer–supplier–supplier relationships. *Journal of Supply Chain Management*, 45(1):8–25.

Chopra, S. and Sodhi, M. (2014). Reducing the risk of supply chain disruptions. *MIT Sloan Management Review*, 55(3):72–80.

Chou, C. (2018). Cainiao to launch smart logistics parks at scale in China. Published on May 31, 2018. Accessed on April 30, 2019. https://www.alizila.com/cainiao-logistics-future-park-china/.

Christopher, M. and Peck, H. (2004). Building the resilient supply chain. *The international journal of logistics management*, 15(2):1–14.

Company, T. B. C. I. B. (2018). The 10th BCI supply chain resilience report. Published on November 6, 2018, accessed on February 26, 2019. https://www.thebci.org/uploads/assets/uploaded/c50072bf-df5c-4c98-a5e1876aafb15bd0.pdf.

Cormican, K. J., Morton, D. P., and Wood, R. K. (1998). Stochastic network interdiction. *Operations Research*, 46(2):184–197.

Cragg, J. G. and Donald, S. G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, 9(2):222–240.

Craighead, C. W., Blackhurst, J., Rungtusanatham, M. J., and Handfield, R. B. (2007). The severity of supply chain disruptions: design characteristics and mitigation capabilities. *Decision Sciences*, 38(1):131–156.

Croissant, Y. and Millo, G. (2018). *Panel Data Econometrics with R*. John Wiley & Sons.

Daft, R. L. and Macintosh, N. B. (1981). A tentative exploration into the amount and equivocality of information processing in organizational work units. *Administrative science quarterly*, 26(2):207–224.

Darr, E. D., Argote, L., and Epple, D. (1995). The acquisition, transfer, and depreciation of knowledge in service organizations: Productivity in franchises. *Management science*, 41(11):1750–1762.

Demeter, K. and Matyusz, Z. (2011). The impact of lean practices on inventory turnover. *International Journal of Production Economics*, 133(1):154–163.

Dong, M. C., Liu, Z., Yu, Y., and Zheng, J.-H. (2015). Opportunism in distribution networks: The role of network embeddedness and dependence. *Production and Operations Management*, 24(10):1657–1670.

Dutton, J. M. and Thomas, A. (1984). Treating progress functions as a managerial opportunity. *Academy of management review*, 9(2):235–247.

Ecklund, D. K. (2010). Warehousing efficiency and effectiveness in the supply chain process. *Supply Chain Management Review*, 15.

Fiksel, J., Polyviou, M., Croxton, K. L., and Pettit, T. J. (2015). From risk to resilience: Learning to deal with disruption. *MIT Sloan Management Review*, 56(2):79–86.

Fisher, M. L., Hammond, J. H., Obermeyer, W. R., and Raman, A. (1994). Making supply meet demand in an uncertain world. *Harvard business review*, 72:83–83.

Fisher, M. L. and Ittner, C. D. (1999). The impact of product variety on automobile assembly operations: Empirical evidence and simulation analysis. *Management science*, 45(6):771–786.

Garson, G. D. (2012). *Hierarchical linear modeling: Guide and applications*. Sage.

Gaur, V., Fisher, M. L., and Raman, A. (2005). An econometric analysis of inventory turnover performance in retail services. *Management science*, 51(2):181–194.

Gaur, V. and Kesavan, S. (2008). The effects of firm size and sales growth rate on inventory turnover performance in the u. sretail sector. In *Retail Supply Chain Management*, pages 25–52. Springer.

Gaur, V., Kesavan, S., Raman, A., and Fisher, M. L. (2007). Estimating demand uncertainty using judgmental forecasts. *Manufacturing & Service Operations Management*, 9(4):480–491.

Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica: Journal of the Econometric Society*, 46(6):1293–1301.

Goldberg, A. V. and Tarjan, R. E. (2014). Efficient maximum flow algorithms. *Communications of the ACM*, 57(8):82–89.

Golden, B. (1978). A problem in network interdiction. *Naval Research Logistics Quarterly*, 25(4):711–713.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American journal of sociology*, 91(3):481–510.

Granovetter, M. (2005). The impact of social structure on economic outcomes. *Journal of economic perspectives*, 19(1):33–50.

Gulati, R. (1998). Alliances and networks. *Strategic management journal*, 19(4):293–317.

Gupta, S. and Roth, A. V. (2007). Martin k. starr: A visionary proponent for system integration, modular production, and catastrophe avoidance. *Production and Operations Management*, 16(1):1–12.

Håkansson, H., Havila, V., and Pedersen, A.-C. (1999). Learning in networks. *Industrial marketing management*, 28(5):443–452.

Hançerlioğulları, G., Şen, A., and Aktunç, E. A. (2016). Demand uncertainty and inventory turnover performance: An empirical analysis of the us retail industry. *International Journal of Physical Distribution & Logistics Management*, 46(6/7):681–708.

Handfield, R. B., Krause, D. R., Scannell, T. V., and Monczka, R. M. (2010). Avoid the pitfalls in supplier development. *MIT Sloan Management Review*, 41(2):37–49.

Haunschild, P. R. and Sullivan, B. N. (2002). Learning from complexity: Effects of prior accidents and incidents on airlines' learning. *Administrative science quarterly*, 47(4):609–643.

Hearnshaw, E. J. and Wilson, M. M. (2013). A complex network approach to supply chain network theory. *International Journal of Operations & Production Management*, 33(4):442–469.

Heikkilä, J. (2002). From supply to demand chain management: efficiency and customer satisfaction. *Journal of operations management*, 20(6):747–767.

Hendricks, K. B. and Singhal, V. R. (2005). 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.

Ho, T.-H., Lim, N., Reza, S., and Xia, X. (2017). Om forum—causal inference models in operations management. *Manufacturing & Service Operations Management*, 19(4):509–525.

Ho, T.-H. and Tang, C. S. (1998). *Product variety management: research advances*, volume 10. Springer Science & Business Media.

Holmes, T. J. (2011). The diffusion of wal-mart and economies of density. *Econometrica*, 79(1):253–302.

Hora, M. and Klassen, R. D. (2013). Learning from others' misfortune: Factors influencing knowledge acquisition to reduce operational risk. *Journal of Operations Management*, 31(1-2):52–61.

Houde, J.-F., Newberry, P., and Seim, K. (2017). Economies of density in e-commerce: A study of amazon's fulfillment center network. Technical report, National Bureau of Economic Research.

Ivanov, D. (2017). Simulation-based ripple effect modelling in the supply chain. *International Journal of Production Research*, 55(7):2083–2101.

Janjarassuk, U. and Nakrachata-Amon, T. (2015). A simulated annealing algorithm to the stochastic network interdiction problem. In *Industrial Engineering and Engineering Management (IEEM), 2015 IEEE International Conference on*, pages 230–233. IEEE.

Jüttner, U. and Maklan, S. (2011). Supply chain resilience in the global financial crisis: an empirical study. *Supply Chain Management: An International Journal*, 16(4):246–259.

Kao, T.-W. D., Simpson, N., Shao, B. B., and Lin, W. T. (2017). Relating supply network structure to productive efficiency: A multi-stage empirical investigation. *European Journal of Operational Research*, 259(2):469–485.

Karp, R. M. (1972). Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer.

Kim, D.-Y. (2014). Understanding supplier structural embeddedness: A social network perspective. *Journal of Operations Management*, 32(5):219–231.

Kim, S.-H. and Tomlin, B. (2013). Guilt by association: Strategic failure prevention and recovery capacity investments. *Management Science*, 59(7):1631–1649.

Kim, Y., Chen, Y.-S., and 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., and Dooley, K. (2011). Structural investigation of supply networks: A social network analysis approach. *Journal of Operations Management*, 29(3):194–211.

Kleindorfer, P. R. and Saad, G. H. (2005). Managing disruption risks in supply chains. *Production and operations management*, 14(1):53–68.

Kolias, G. D., Dimelis, S. P., and Filios, V. P. (2011). An empirical analysis of inventory turnover behaviour in greek retail sector: 2000–2005. *International Journal of Production Economics*, 133(1):143–153.

Kutner, M. H., Nachtsheim, C., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models*. McGraw-Hill/Irwin, New York.

Ladyman, J., Lambert, J., and Wiesner, K. (2013). What is a complex system? *European Journal for Philosophy of Science*, 3(1):33–67.

Lancaster, K. (1990). The economics of product variety: A survey. *Marketing science*, 9(3):189–206.

Lapré, M. A., Mukherjee, A. S., and Van Wassenhove, L. N. (2000). Behind the learning curve: Linking learning activities to waste reduction. *Management Science*, 46(5):597–611.

Lapré, M. A., Nembhard, I. M., et al. (2011). Inside the organizational learning curve: Understanding the organizational learning process. *Foundations and Trends® in Technology, Information and Operations Management*, 4(1):1–103.

Lawson, B., Tyler, B. B., and Cousins, P. D. (2008). Antecedents and consequences of social capital on buyer performance improvement. *Journal of Operations management*, 26(3):446–460.

Lee, H.-H., Zhou, J., and Hsu, P.-H. (2015). The role of innovation in inventory turnover performance. *Decision Support Systems*, 76:35–44.

Lee, H. L. (2004). The triple-A supply chain. *Harvard business review*, 82(10):102–113.

Lee, H. L. and Billington, C. (1993). Material management in decentralized supply chains. *Operations research*, 41(5):835–847.

Lee, H. L., Padmanabhan, V., and Whang, S. (1997a). The bullwhip effect in supply chains. *Sloan management review*, 38:93–102.

Lee, H. L., Padmanabhan, V., and Whang, S. (1997b). Information distortion in a supply chain: The bullwhip effect. *Management science*, 43(4):546–558.

Levinthal, D. A. (1997). Adaptation on rugged landscapes. *Management science*, 43(7):934–950.

Levy, F. K. (1965). Adaptation in the production process. *Management Science*, 11(6):B–136.

Liker, J. K. and Choi, T. Y. (2004). Building deep supplier relationships. *Harvard business review*, 82(12):104–113.

Lin, J. L., Fang, S.-C., Fang, S.-R., and Tsai, F.-S. (2009). Network embeddedness and technology transfer performance in r&d consortia in taiwan. *Technovation*, 29(11):763–774.

Lokam, S. V. (2003). Graph complexity and slice functions. *Theory of Computing Systems*, 36(1):71–88.

Madsen, P. M. (2009). These lives will not be lost in vain: Organizational learning from disaster in us coal mining. *Organization Science*, 20(5):861–875.

Madsen, P. M. and Desai, V. (2010). Failing to learn? the effects of failure and success on organizational learning in the global orbital launch vehicle industry. *Academy of Management Journal*, 53(3):451–476.

Madsen, P. M. and Desai, V. (2018). No firm is an island: The role of population-level actors in organizational learning from failure. *Organization Science*, 29(4):739–753.

Magnanti, T. L., Shen, Z.-J. M., Shu, J., Simchi-Levi, D., and Teo, C.-P. (2006). Inventory placement in acyclic supply chain networks. *Operations Research Letters*, 34(2):228–238.

Mapes, J. (2015). Inventory performance measures. *Wiley Encyclopedia of Management*, 10:1.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the royal statistical society. Series B (Methodological)*, 42(2):109–142.

Medal, H. R., Pohl, E. A., and Rossetti, M. D. (2016). Allocating protection resources to facilities when the effect of protection is uncertain. *IIE Transactions*, 48(3):220–234.

Melnyk, S. A., Closs, D. J., Griffis, S. E., Zobel, C. W., and Macdonald, J. R. (2015). Understanding supply chain resilience. Published on November 20, 2015, accessed on February 14, 2019. https://www.supplychain247.com/article/understanding_supply_chain_resilience.

Mochari, I. (2016). Why half of the S&P 500 companies will be replaced in the next decade. Published on March 23, 2016, accessed on April 5, 2019. https://www.inc.com/ilan-mochari/innosight-sp-500-new-companies.html.

Moran, P. (2005). Structural vs. relational embeddedness: Social capital and managerial performance. *Strategic management journal*, 26(12):1129–1151.

Muller, J. (2018). How ford moved heaven and earth to save f-series production (and profits) after devastating fire. Published on May 16, 2018, accessed on April 5, 2019. https://www.forbes.com/sites/joannmuller/2018/05/16/how-ford-moved-heaven-and-earth-to-save-f-series-production-and-profits-after-devastating-fire/.

Nagamochi, H. and Ibaraki, T. (1991). Maximum flows in probabilistic networks. *Networks*, 21(6):645–666.

Nahapiet, J. and Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of management review*, 23(2):242–266.

Nahmias, S. (2008). *Production and Operations Analysis*. McGraw-Hill/Irwin, Boston.

Nair, A. and 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.

Norrman, A. and Jansson, U. (2004). Ericsson's proactive supply chain risk management approach after a serious sub-supplier accident. *International journal of physical distribution & logistics management*, 34(5):434–456.

Peeta, S., Salman, F. S., Gunnec, D., and Viswanath, K. (2010). Pre-disaster investment decisions for strengthening a highway network. *Computers & Operations Research*, 37(10):1708–1719.

Pettit, T. J., Croxton, K. L., and Fiksel, J. (2013). Ensuring supply chain resilience: development and implementation of an assessment tool. *Journal of business logistics*, 34(1):46–76.

Podolny, J. M. and Baron, J. N. (1997). Resources and relationships: Social networks and mobility in the workplace. *American sociological review*, 62(5):673–693.

Ponomarov, S. Y. and Holcomb, M. C. (2009). Understanding the concept of supply chain resilience. *The international journal of logistics management*, 20(1):124–143.

Pudlák, P. and Rödl, V. (1994). Some combinatorial-algebraic problems from complexity theory. *Discrete Mathematics*, 136(1-3):253–279.

Rabinovich, E., Dresner, M. E., and Evers, P. T. (2003). Assessing the effects of operational processes and information systems on inventory performance. *Journal of Operations Management*, 21(1):63–80.

Rajagopalan, S. (2013). Impact of variety and distribution system characteristics on inventory levels at us retailers. *Manufacturing & Service Operations Management*, 15(2):191–204.

Ramdas, K. (2003). Managing product variety: An integrative review and research directions. *Production and operations management*, 12(1):79–101.

Rumyantsev, S. and Netessine, S. (2007). What can be learned from classical inventory models? a cross-industry exploratory investigation. *Manufacturing & Service Operations Management*, 9(4):409–429.

Sáenz, M. J. and Revilla, E. (2013). Cisco Systems, Inc.: supply chain risk management. *The Supply Chain Management Casebook: Comprehensive Coverage and Best Practices in SCM, Financial Times Press, Upper Saddle River, NJ*, pages 80–95.

Salmeron, J., Wood, K., and Baldick, R. (2004). Analysis of electric grid security under terrorist threat. *IEEE Transactions on power systems*, 19(2):905–912.

Sanderson, E. and Windmeijer, F. (2016). A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of Econometrics*, 190(2):212–221.

Scheibe, K. P. and 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.

Schmenner, R. W. (2001). Looking ahead by looking back: Swift, even flow in the history of manufacturing. *Production and Operations Management*, 10(1):87–96.

Schmenner, R. W. and Swink, M. L. (1998). On theory in operations management. *Journal of operations management*, 17(1):97–113.

Shannon, C. E. and Weaver, W. (1949). *The Mathematical Theory of Communication.* Urbana: University of Illinois Press, Champaign, IL.

Sharma, M. and Ghosh, D. (2013). Speeding up the estimation of the expected value of maximum flow through reliable networks. *IEEE Transactions on Reliability*, 62(1):105–115.

Sheffi, Y. and Rice Jr, J. B. (2005). A supply chain view of the resilient enterprise. *MIT Sloan management review*, 47(1):41.

Silver, D. (2016). The automotive supply chain, explained. Published on May 30, 2016, accessed on September 23, 2018. https://medium.com/self-driving-cars/the-automotive-supply-chain-explained-d4e74250106f.

Simchi-Levi, D. (2013). OM forum – OM research: From problem-driven to data-driven research. *Manufacturing & Service Operations Management*, 16(1):2–10.

Simchi-Levi, D., Schmidt, W., and Wei, Y. (2014). From superstorms to factory fires: Managing unpredictable supply chain disruptions. *Harvard Business Review*, 92(1-2):96–101.

Simchi-Levi, D., Schmidt, W., Wei, Y., Zhang, P. Y., Combs, K., Ge, Y., Gusikhin, O., Sanders, M., and Zhang, D. (2015). Identifying risks and mitigating disruptions in the automotive supply chain. *Interfaces*, 45(5):375–390.

Smith, S. A. and Agrawal, N. (2000). Management of multi-item retail inventory systems with demand substitution. *Operations Research*, 48(1):50–64.

Smunt, T. L. and Ghose, S. (2016). An entropy measure of flow dominance for predicting operations performance. *Production and Operations Management*, 25(10):1638–1657.

Snyder, L. V. and Daskin, M. S. (2007). Models for reliable supply chain network design. In *Critical infrastructure*, pages 257–289. Springer.

Sribney, W., Wiggins, V., and Drukker, D. (2018). Negative and missing R-squared for 2SLS/IV. Accessed on October 22, 2018. https://www.stata.com/support/faqs/statistics/two-stage-least-squares/.

Stampler, L. (2019). It's not just Amazon: Free one-day delivery is the new normal. Published on April 30, 2019. Accessed on April 30, 2019. http://fortune.com/2019/04/30/walmart-target-and-amazon-prime-one-day-shipping/.

Starr, M. K. (1980). Some new fundamental considerations of variety-seeking behavior. *Behavioral Science*, 25(3):171–179.

Statista (2018). Selected automakers' U.S. YTD market share in september 2018. Published in September 2018, accessed on February 27, 2019. https://www.statista.com/statistics/343162/market-share-of-major-car-manufacturers-in-the-united-states/.

Staudt, F. H., Alpan, G., Di Mascolo, M., and Rodriguez, C. M. T. (2015). Warehouse performance measurement: a literature review. *International Journal of Production Research*, 53(18):5524–5544.

Stock, J. H. and Watson, M. W. (2011). *Introduction to Econometrics*. Pearson.

Straathof, S. M. (2007). Shannon's entropy as an index of product variety. *Economics Letters*, 94(2):297–303.

Sun, H. and Wu, J. (2005). Scale-free characteristics of supply chain distribution networks. *Modern physics letters B*, 19(17):841–848.

Tabachnick, B. G., Fidell, L. S., and Ullman, J. B. (2007). *Using multivariate statistics*, volume 5. Pearson Boston, MA.

Tang, C. S. (2006). Robust strategies for mitigating supply chain disruptions. *International Journal of Logistics: Research and Applications*, 9(1):33–45.

Tomlin, B. (2006). On the value of mitigation and contingency strategies for managing supply chain disruption risks. *Management Science*, 52(5):639–657.

Ton, Z. and Raman, A. (2010). The effect of product variety and inventory levels on retail store sales: A longitudinal study. *Production and Operations Management*, 19(5):546–560.

Uzzi, B. (1996). The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American sociological review*, 61(4):674–698.

Vogelsang, T. J. (2012). Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. *Journal of Econometrics*, 166(2):303–319.

Wallace, S. W. (1987). Investing in arcs in a network to maximize the expected max flow. *Networks*, 17(1):87–103.

Wan, X., Evers, P. T., and Dresner, M. E. (2012). Too much of a good thing: The impact of product variety on operations and sales performance. *Journal of Operations Management*, 30(4):316–324.

Wang, Y., Gilland, W., and Tomlin, B. (2010). Mitigating supply risk: Dual sourcing or process improvement? *Manufacturing & Service Operations Management*, 12(3):489–510.

Wang, Y., Xiao, Y., and Yang, N. (2014). Improving reliability of a shared supplier with competition and spillovers. *European Journal of Operational Research*, 236(2):499–510.

Wang, Y. I., Li, J., and Anupindi, R. (2017). Risky suppliers or risky supply chains? an empirical analysis of sub-tier supply network structure on firm risk in the high-tech sector.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of ”small-world” networks. *nature*, 393(6684):440.

Wollmer, R. D. (1991). Investments in stochastic maximum flow networks. *Annals of Operations Research*, 31(1):457–467.

Wu, T., Blackhurst, J., and O'grady, P. (2007). Methodology for supply chain disruption analysis. *International journal of production research*, 45(7):1665–1682.

Wu, Z. and Choi, T. Y. (2005). Supplier–supplier relationships in the buyer–supplier triad: Building theories from eight case studies. *Journal of Operations management*, 24(1):27–52.

Wu, Z., Choi, T. Y., and Rungtusanatham, M. J. (2010). Supplier–supplier relationships in buyer–supplier–supplier triads: Implications for supplier performance. *Journal of Operations Management*, 28(2):115–123.

Yelle, L. E. (1979). The learning curve: Historical review and comprehensive survey. *Decision sciences*, 10(2):302–328.

Zhao, K., Kumar, A., Harrison, T. P., and Yen, J. (2011). Analyzing the resilience of complex supply network topologies against random and targeted disruptions. *IEEE Systems Journal*, 5(1):28–39.

Zhou, Y. M. and Wan, X. (2017a). Product variety and vertical integration. *Strategic Management Journal*, 38(5):1134–1150.

Zhou, Y. M. and Wan, X. (2017b). Product variety, sourcing complexity, and the bottleneck of coordination. *Strategic Management Journal*, 38(8):1569–1587.

Zipkin, P. H. (2000). *Foundations of inventory management*.

Zobel, C. W. (2014). Quantitatively representing nonlinear disaster recovery. *Decision Sciences*, 45(6):1053–1082.

Zobel, C. W. and Khansa, L. (2012). Quantifying cyberinfrastructure resilience against multi-event attacks. *Decision Sciences*, 43(4):687–710.

Appendix A

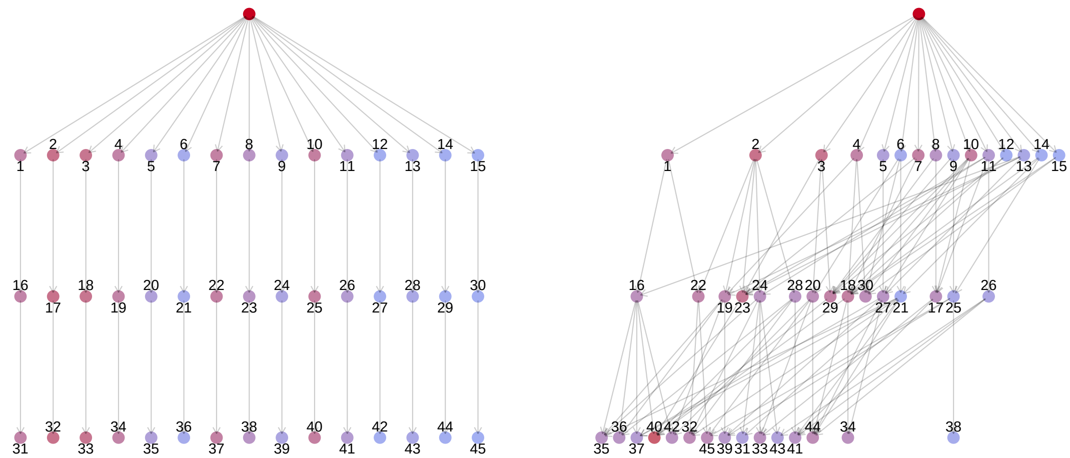
Essay 2 Appendices

A.1 Robustness Check on Derived Networks

To examine the robustness of our data-driven prescription, we create two networks based on Company A's realistic distribution network ("DN"): one with a linear structure ("AN1", Figure A.1(a)) and the other with a random tiered structure ("AN2", Figure A.1(b)). The two derived networks have the same number of nodes as DN and both have three layers. The first 15 nodes in both AN1 and AN2 have the same capacities as nodes 1 through 15 in DN. Other nodes' capacities equal the sum of incoming edge capacities. Edge capacity is the emanating node's capacity divided by its out-degree. Table A.1 shows the node information of AN2.

Table A.2 displays the optimal investment for the linear-structured AN1. There is no routing flexibility in AN1. The investment tends to complete entire paths under any combination of p^R and δp^D. The path with the highest-capacitated node (i.e., node 2) is invested first, followed by the path with the next (fourth, in this context) highest-capacitated node, and so on.

Table A.4 displays the optimal investment for AN2 under nonreroutable and reroutable flows. As the main analysis, we perform ordered logistic regression as in Table A.3. The interaction between p^D and node capacity is significantly positive



(a) AN1 with a linear structure (b) AN2 with a random tiered structure

Figure A.1: Two derived networks based on the distribution network

Table A.1: Node capacity with rank (in brackets) and # paths through node for AN2

Node	Capacity	$\#\phi$	Node	Capacity	$\#\phi$	Node	Capacity	$\#\phi$
1 [10]	0.112	7	16 [18]	0.0569	10	31 [35]	0.0101	4
2 [2]	0.221	16	17 [19]	0.0542	6	32 [14]	0.0712	7
3 [3]	0.197	6	18 [6]	0.130	18	33 [21]	0.0536	7
4 [9]	0.116	8	19 [12]	0.0838	12	34 [20]	0.0537	9
5 [33]	0.0131	4	20 [17]	0.0659	8	35 [26]	0.0424	9
6 [41]	1.04e-03	11	21 [42]	4.10e-04	12	36 [25]	0.0434	6
7 [7]	0.127	4	22 [11]	0.100	4	37 [32]	0.0219	7
8 [27]	0.0418	3	23 [4]	0.180	5	38 [40]	1.10e-03	2
9 [36]	6.48e-03	4	24 [22]	0.0499	10	39 [30]	0.0325	8
10 [5]	0.138	11	25 [39]	3.30e-03	6	40 [1]	0.363	19
11 [31]	0.0228	11	26 [37]	5.70e-03	3	41 [29]	0.0401	7
12 [43]	2.29e-04	4	27 [28]	0.0417	12	42 [23]	0.0488	7
13 [38]	3.99e-03	14	28 [24]	0.0442	4	43 [34]	0.0119	3
14 [44]	1.86e-04	12	29 [8]	0.117	5	44 [13]	0.0721	14
15 [45]	4.58e-05	4	30 [16]	0.0666	4	45 [15]	0.0677	6
Total Paths: 119						Grand Mean:	0.0652	7.844

Table A.2: Optimal investing structure for AN1

K	Any $(p^R, \delta p^D)$ combination
1	2 or 17 or 32
2	17,2 or 32,2 or 32,17
3	32,17,2 (denoted as "ϕ_1")
4	3,ϕ_1 or 18,ϕ_1 or 33,ϕ_1
5	18,3,ϕ_1 or 33,3,ϕ_1 or 33,18,ϕ_1
6	33,18,3 (ϕ_2),ϕ_1
7	10,ϕ_2,ϕ_1 or 25,ϕ_2,ϕ_1 or 40,ϕ_2,ϕ_1
8	25,10,ϕ_2,ϕ_1 or 40,10,ϕ_2,ϕ_1 or 40,25,ϕ_2,ϕ_1
9	40,25,10 (ϕ_3),ϕ_2,ϕ_1
10	7,ϕ_3,ϕ_2,ϕ_1 or 22,ϕ_3,ϕ_2,ϕ_1 or 37,ϕ_3,ϕ_2,ϕ_1

in Models A2 (44.29^{*}) and A5 ($24.564^{\dagger}$) – Hypothesis 5 is partially supported. The main effect of node capacity plays a dominant role, possibly because one or a few paths dominate the capacity (with high path flows) and nodes with high capacities lie on those paths. For instance, paths 2-23-40 and 3-23-40 have highest node capacities and possibly highest path flows, which will very likely be invested under different disruption probabilities. In this sense, the result does not negate our hypothesis, but is more of a special case. For Hypothesis 6, the interaction between p^D and average path length is positive in Models A2 (32.417^{***}), A4 (11.727^{***}), and A5 (12.579^{***}) – Hypothesis 6 is verified. Hypothesis 7 is verified by a negative interaction between routing and flow centrality (-13.833^{***}).

Table A.3: Ordered logistic regression results for AN2

	Response: *First Appearance*				
	Model A1	Model A2	Model A3	Model A4	Model A5
p^D	-0.679	-103.972^{***}	$-17.372^{\dagger}$	-38.099^{***}	-41.146^{***}
	(0.637)	(28.257)	(9.234)	(9.706)	(10.307)
δ	0.044	-0.878	0.859	-0.041	-0.149
	(0.506)	(0.749)	(0.723)	(0.514)	(0.512)
rout	0.063			0.059	2.253^{***}
	(0.275)			(0.275)	(0.587)
ω_i	-37.321^{***}	-72.693^{***}	-48.394^{**}	-51.909^{***}	-56.614^{***}
	(3.737)	(16.303)	(14.783)	(9.936)	(10.187)
$\overline{\lvert k\rvert}_i$	2.233^{***}	-18.936^{**}	-0.612	-4.803^{*}	-5.194^{*}
	(0.622)	(7.186)	(2.313)	(2.272)	(2.469)
$C_\phi(i)$	1.486	9.156^{***}	-6.858	1.557	8.410^{***}
	(1.684)	(2.601)	(6.924)	(1.711)	(2.332)
$p^D \times \omega_i$		44.290^{*}	4.599	20.594	$24.564^{\dagger}$
		(21.661)	(21.658)	(14.033)	(14.187)
$p^D \times \overline{\lvert k\rvert}_i$		32.417^{***}	5.859	11.727^{***}	12.579^{***}
		(9.741)	(3.664)	(3.488)	(3.672)
rout $\times C_\phi(i)$					-13.833^{***}
					(3.258)
N	184	92	92	184	184
logLik	-335.093	-147.931	-151.127	-322.778	-313.445
AIC	700.19	327.86	336.25	679.56	662.89

Note: †p $<$ .1; *p<0.05; **p<0.01; ***p<0.001

Table A.4: Optimal investment for AN2 as K increases from 1 to 10

Different $(p^R, \delta p^D)$ combinations, under nonreroutable flow					
K	$(0.2, 0.1)$	$(0.4, 0.1)$	$(0.6, 0.1)$	$(0.8, 0.1)$	
1	40	40	40	40	
2	4,30	2,40	2,40	2,40	
3	4,30,40	3,23,40	2,3,40	2,3,40	
4	4,10,30,40	2,3,23,40	2,3,23,40	2,3,23,40	
5	4,10,23,30,40	2,3,10,23,40	2,3,10,23,40	2,3,10,23,40	
6	3,4,10,23,30,40	2,3,10,23,30,40	2,3,10,23,29,40	2,3,7,10,23,40	
7	2,3,4,10,23,30,40	2,3,4,10,23,30,40	2,3,7,10,23,29,40	2,3,7,10,18,23,40	
8	2,3,4,10,23,29,30,40	2,3,4,10,23,29,30,40	2,3,7,10,18,23,29,40	2,3,7,10,18,23,29,40	
9	2,3,4,7,10,23,29,30,40	2,3,4,7,10,23,29,30,40	2,3,4,7,10,18,23,29,40	2,3,4,7,10,18,23,29,40	
10	2,3,4,7,10,18,23,29,30,40	2,3,4,7,10,18,23,29,30,40	2,3,4,7,10,18,23,29,30,40	2,3,4,7,10,18,19,23,29,40	
K	$(0.2, 0.3)$	$(0.2, 0.5)$	$(0.2, 0.7)$	$(0.4, 0.5)$	$(0.6, 0.3)$
1	40	40	40	40	40
2	4,30	4,30	4,30	23,40	23,40
3	3,23,40	3,23,40	3,23,40	3,23,40	3,23,40
4	3,23,29,40	3,23,29,40	3,23,29,40	3,23,29,40	2,3,23,40
5	3,7,23,29,40	3,7,23,29,40	3,7,23,29,40	3,7,23,29,40	2,3,7,23,40
6	2,3,7,23,29,40	2,3,7,23,29,40	2,3,7,23,29,40	2,3,7,23,29,40	2,3,7,23,29,40
7	3,4,10,23,29,30,40	2,3,7,18,23,29,40	2,3,7,18,23,29,40	2,3,7,18,23,29,40	2,3,7,18,23,29,40
8	2,3,4,10,23,29,30,40	2,3,7,10,18,23,29,40	2,3,7,10,18,23,29,40	2,3,7,10,18,23,29,40	2,3,7,10,18,23,29,40
9	3,4,7,10,18,23,29,30,40	3,4,7,10,18,23,29,30,40	2,3,7,10,18,23,29,40,44	2,3,7,10,18,23,29,40,44	2,3,4,7,10,18,23,29,40
10	2,3,4,7,10,18,23,29,30,40	2,3,4,7,10,18,23,29,30,40	2,3,4,7,10,18,23,29,30,40	2,3,4,7,10,18,23,29,30,40	2,3,4,7,10,18,23,29,30,40
Different $(p^R, \delta p^D)$ combinations, under reroutable flow					
K	$(0.2, 0.1)$	$(0.4, 0.1)$	$(0.6, 0.1)$	$(0.8, 0.1)$	
1	40	40	40	40	
2	23,40	2,40	2,40	2,40	
3	2,23,40	2,30,40	2,18,40	3,23,40	
4	2,3,23,40	2,18,30,40	2,10,18,40	2,3,23,40	
5	2,3,18,23,40	2,10,18,30,40	2,10,18,23,40	2,3,18,23,40	
6	2,10,18,23,30,40	2,18,19,23,30,40	2,3,18,19,23,40	2,3,18,23,29,40	
7	2,3,4,18,23,30,40	2,4,10,18,23,30,40	2,3,10,18,23,30,40	2,3,7,18,23,29,40	
8	2,4,10,18,23,29,30,40	2,4,7,18,19,23,30,40	2,3,4,18,19,23,29,40	2,3,7,10,18,23,29,40	
9	2,3,4,7,10,18,23,30,40	2,4,10,18,19,23,29,30,40	2,3,7,10,18,23,29,30,40	2,3,7,10,18,19,23,29,40	
10	2,3,4,7,10,18,23,29,30,40	2,3,4,7,10,18,23,29,30,40	2,3,4,10,18,19,23,29,30,40	2,3,7,10,18,19,23,29,30,40	
K	$(0.2, 0.3)$	$(0.2, 0.5)$	$(0.2, 0.7)$	$(0.4, 0.5)$	$(0.6, 0.3)$
1	40	40	40	40	40
2	2,40	23,40	23,40	23,40	23,40
3	2,23,40	2,23,40	2,23,40	2,23,40	3,23,40
4	2,23,30,40	2,18,23,40	2,18,23,40	2,3,23,40	2,3,23,40
5	2,4,23,30,40	2,10,18,23,40	2,3,23,29,40	2,3,23,29,40	2,3,18,23,40
6	2,4,18,23,30,40	2,3,18,23,29,40	2,3,18,23,29,40	2,3,18,23,29,40	2,3,18,23,29,40
7	2,4,10,18,23,30,40	2,3,4,18,23,29,40	2,3,10,18,23,29,40	2,3,7,18,23,29,40	2,3,10,18,23,29,40
8	2,4,10,18,23,29,30,40	2,3,4,18,23,29,30,40	2,3,10,18,23,29,30,40	2,3,7,18,23,29,30,40	2,3,10,18,23,29,30,40
9	2,3,4,10,18,23,29,30,40	2,3,4,10,18,23,29,30,40	2,3,10,18,20,23,29,30,40	2,3,7,10,18,23,29,30,40	2,3,10,18,22,23,29,30,40
10	2,3,4,10,18,19,23,29,30,40	2,3,4,10,18,19,23,29,30,40	2,3,10,18,20,23,29,30,40,44	2,3,10,18,22,23,29,30,32,40	2,3,4,10,18,22,23,29,30,40

Appendix B

Essay 3 Appendices

B.1 Proof

Proof B.1

Proof of Lemma 2. The statement holds for the source nodes, as shown in Equation (4.13). For the non-source non-sink nodes, we use complete induction to prove the statement. To facilitate the proof, we classify the nodes based on the acyclic structure. Nodes without any incoming edges are the *source nodes* and are classified as "level" 0. Nodes with incoming edges *only* from level-0 nodes are at level 1. Nodes that have incoming edges *only* from level-0 and 1 nodes (i.e., lower-level nodes) are at level 2, and so on, till there is only the sink in the network. Let $V^n, 0 \leq n \leq M$ be the set of level-n nodes, where M denotes the largest number of the levels.

First, we prove the base case that the limiting distribution exists for any level-1 node $i \in V^1$. As time goes to infinity, we obtain the limiting transition probabilities for i as

$$
\begin{aligned}
\lim_{t\to\infty}\beta_i^t &= 1-(1-\beta[\lim_{t\to\infty} CumDis_i^t+1]^{-a_i})\prod_{j|(j,i)\in E}(1-\alpha\pi_j^D) \\
&= \begin{cases} 1-\prod_{j|(j,i)\in E}(1-\alpha\pi_j^D), & \text{for } a_i>0, \forall i\in V^1 \\ 1-(1-\beta)\prod_{j|(j,i)\in E}(1-\alpha\pi_j^D), & \text{for } a_i=0, \forall i\in V^1 \end{cases}
\end{aligned} \tag{B.1}
$$

$$
\begin{aligned}
\lim_{t\to\infty}\gamma_i^t &= [1-(1-\gamma)e^{-b_i\lim_{t\to\infty} CumDis_i^t}]\prod_{j|(j,i)\in E}(1-\alpha\pi_j^D) \\
&= \begin{cases} \prod_{j|(j,i)\in E}(1-\alpha\pi_j^D), & \text{for } b_i>0, \forall i\in V^1 \\ \gamma\prod_{j|(j,i)\in E}(1-\alpha\pi_j^D), & \text{for } b_i=0, \forall i\in V^1 \end{cases}
\end{aligned} \tag{B.2}
$$

where π_j^D is the limiting probability of disruption for i's supplier j (j is a source node). From Equation (4.13), π_j^D is a constant no matter how j learns. Therefore, $\prod_{j|(j,i)\in E}(1-\alpha\pi_j^D)$ is a constant. The limiting transition probabilities hence are constant numbers for all combinations of a_i and b_i. The limiting distribution exists for any level-1 node and takes exactly the form of Equation (4.14) (by plugging in $\lim_{t\to\infty}\beta_i^t$ and $\lim_{t\to\infty}\gamma_i^t$ from Equations (B.1) and (B.2) to (4.14)).

Next, we consider the step case and prove the statement for level-$(n+1)$ nodes under the assumption that the statement holds for levels lower than $n+1$ ($0\leq n\leq M-1$). The reason to adopt complete induction is that a higher-level node may be supplied by nodes at different lower levels (think of a simple example of the triad structure). Let i be an arbitrary node at level-$(n+1)$. As time goes to infinity, the limiting transition probabilities for $i\in V^{n+1}$ takes the form of Equations (B.1) and (B.2) (with V^1 changed to V^{n+1}) where j is still i's supplier but now may be at any level lower than $n+1$. Still, π_j^D and $\prod_{j|(j,i)\in E}(1-\alpha\pi_j^D)$ are constant numbers no matter how j learns. Therefore, the limiting transition probabilities are constant for all combinations of a_i and b_i, so the limiting distribution exists for any $i\in V^{n+1}$ in the same form of Equation (4.14). In general, with the complete induction, Lemma

2 is demonstrated. □

Proof B.2

Proof of Proposition 5. Based on Lemma 2, the probability of a node's healthy state is well-defined at infinite time. Given a specific scenario with respect to which nodes are disrupted, the corresponding PDP can be computed. Particularly, the probability of a disrupted path is the probability of at least one disrupted node on the path. Hence, for any path $\phi \in \Phi$, the infinite-time probability of disruption is $1 - \prod_{i \in \phi} \pi_i^H$.

We sum up the probabilities with respect to all paths and divide it by the number of paths ($|\Phi|$) to get the expected PDP$^\infty$.

$$\mathbb{E}[\mathrm{PDP}^\infty] \overset{\text{def}}{=} \mathbb{E}[\lim_{t\to\infty} \mathrm{PDP}^t] = \frac{\sum_{\phi\in\Phi} (1 - \prod_{i\in\phi} \pi_i^H)}{|\Phi|} = 1 - |\Phi|^{-1} \sum_{\phi\in\Phi} \prod_{i\in\phi} \pi_i^H \quad \text{(B.3)}$$

Hence, Proposition 5 is demonstrated. □

B.2 Additional Visualization

We include all the visualization of the mean PDP trajectories across all networks in Experiment I, II, and *post hoc* analysis in a separate zip file (available upon request). For Experiment I, one network has two files ending with "viz1" and "viz2", e.g., "Honda_Experiment_I_viz1.pdf". The two files reflect the same mean PDP trajectories, but differ in grouping factors – "viz1" groups the trajectories by risk ("MC_set" in the figure, where the three numbers stand for α, β, and γ respectively) and learning-to-recover, whereas "viz2" groups the trajectories by risk and learning-to-prevent.

For Experiment II and *post hoc* analysis, each network has one associated file, e.g., "Honda_Experiment_II.pdf" and "Honda_posthoc.pdf". For the two files of visualization of the same network, the "benchmark" curves indicate suppliers' adoption of the corresponding same rates (a, b) (data from Experiment I), and other legends are self-explanatory. The three numbers for “risk” stand for α, β, and γ respectively.

B.3 Robustness Check: Simulation Length of 500

As indicated in the main text, we reduce the simulaiton length from 1000 to 500 as the robustness check. In this sense, the area under curve (AUC) is calculated through PDP values from time 1 till time 500. Except for the difference in simulation length, other factors and the analysis in the robustness check are the same as in the main analysis.

For Experiment I, Table B.1 displays the robustness check results, corresponding to Table 4.3. We also visualize the interaction effects in Figure B.1, which corresponds to Figure 4.6. Through examination, it is apparent that all the coefficients share the same direction and significance with their counterparts in the main analysis. Moreover, the values and the significance magnitude of the coefficients are very similar to the main analysis. Figure B.1 exhibits very similar interaction effects as well. In this sense, we conclude the robustness of our results for Experiment I.

For Experiment II, Table B.2 displays the robustness check results, corresponding to Table 4.4. We also visualize the interaction effects in Figure B.2, which corresponds to Figure 4.7. Most coefficients share the same direction and significance with their counterparts in the main analysis. Their values and the significance magnitude are similar between the main and robustness results. Figure B.2 exhibits very similar interaction effects as well. In this sense, we conclude the robustness of our results for Experiment II.

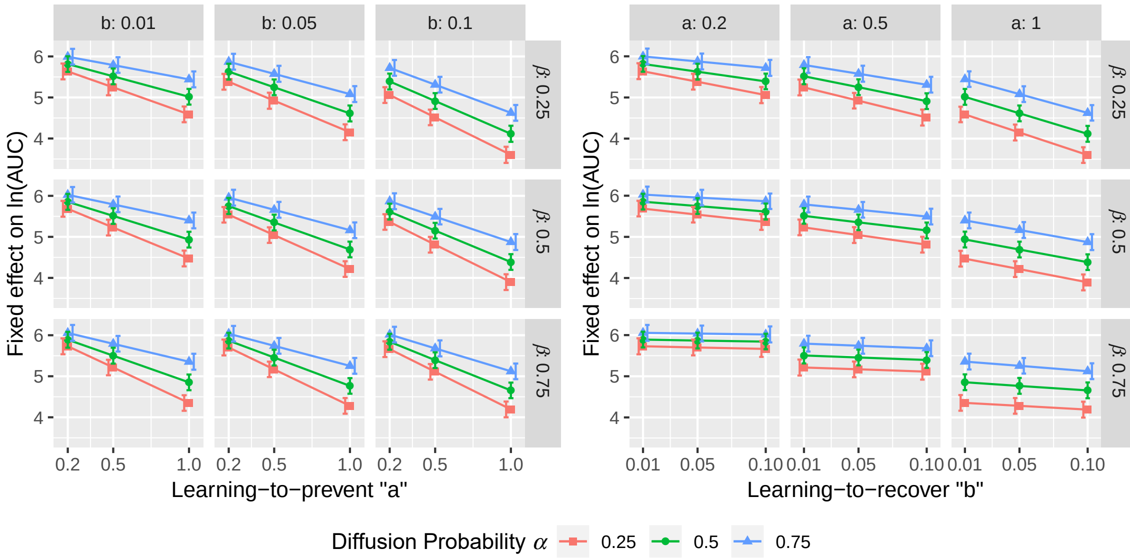


Figure B.1: Interaction effects between risk and supplier learning in Experiment I (simulation length 500)

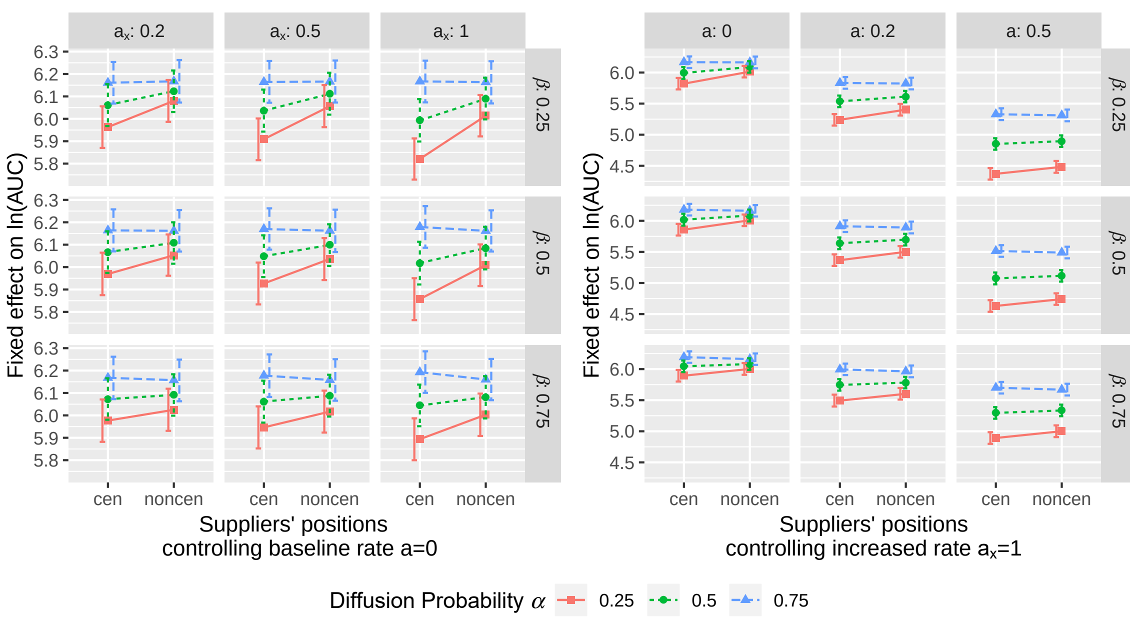


Figure B.2: Interaction effects between centrality, risk, and supplier learning in Experiment II (simulation length 500)

Table B.1: Robustness check on Experiment I (same supplier learning rates) – simulation length 500

	Mixed-Effects			*OLS* (robustness check)			
DV: ln(AUC)	(R1) All	(R2) All	(R3) All	(R4) Honda	(R5) Toyota	(R6) Rnd.1	(R7) Rnd.2
Constant	5.461***	5.951***	5.779***	5.797***	5.845***	5.746***	5.727***
	(0.034)	(0.034)	(0.036)	(0.019)	(0.019)	(0.027)	(0.026)
Diffusion (α)	1.102***	0.400***	0.441***	0.397***	0.391***	0.496***	0.481***
	(0.006)	(0.009)	(0.024)	(0.036)	(0.035)	(0.049)	(0.048)
Disruption (β)	0.330***	0.052***	0.306***	0.276***	0.267***	0.348***	0.335***
	(0.006)	(0.009)	(0.024)	(0.036)	(0.035)	(0.049)	(0.048)
Prevent (a)	−1.280***	−1.806***	−1.290***	−1.153***	−1.023***	−1.492***	−1.494***
	(0.003)	(0.009)	(0.023)	(0.034)	(0.034)	(0.047)	(0.046)
Recover (b)	−3.636***	−10.286***	−10.348***	−9.007***	−8.269***	−12.139***	−11.977***
	(0.032)	(0.089)	(0.230)	(0.347)	(0.340)	(0.473)	(0.466)
$\alpha \times a$		1.374***	1.056***	0.612***	0.610***	1.507***	1.496***
		(0.013)	(0.042)	(0.063)	(0.062)	(0.087)	(0.085)
$\alpha \times b$		2.933***	10.630***	7.883***	7.968***	13.691***	12.978***
		(0.120)	(0.427)	(0.642)	(0.629)	(0.875)	(0.862)
$\beta \times a$		−0.322***	−1.142***	−0.973***	−1.042***	−1.282***	−1.272***
		(0.013)	(0.042)	(0.063)	(0.062)	(0.087)	(0.085)
$\beta \times b$		10.366***	12.732***	11.095***	10.206***	15.000***	14.627***
		(0.120)	(0.427)	(0.642)	(0.629)	(0.875)	(0.862)
$\alpha \times \beta$			−0.195***	−0.152*	−0.186**	−0.237**	−0.205*
			(0.044)	(0.067)	(0.065)	(0.091)	(0.090)
$a \times b$			−6.777***	−7.710***	−9.335***	−4.928***	−5.136***
			(0.406)	(0.611)	(0.598)	(0.832)	(0.820)
$\alpha \times \beta \times a$			0.902***	0.938***	1.072***	0.802***	0.795***
			(0.078)	(0.118)	(0.115)	(0.160)	(0.158)
$\alpha \times \beta \times b$			−12.575***	−9.309***	−9.457***	−16.360***	−15.174***
			(0.790)	(1.190)	(1.164)	(1.621)	(1.596)
$\alpha \times a \times b$			−4.264***	2.286*	3.870***	−12.093***	−11.121***
			(0.751)	(1.131)	(1.107)	(1.541)	(1.518)
$\beta \times a \times b$			8.280***	8.156***	9.998***	7.192***	7.776***
			(0.751)	(1.131)	(1.107)	(1.541)	(1.518)
$\alpha \times \beta \times a \times b$			1.893	−3.187	−4.831*	8.703**	6.886*
			(1.391)	(2.095)	(2.050)	(2.854)	(2.811)
Observations	28,800	28,800	28,800	7,200	7,200	7,200	7,200
Adjusted R^2				0.960	0.956	0.948	0.949
Log Likelihood	3,385.6	11,152.4	14,564.2				
AIC	−6,757	−22,283	−29,092				
BIC	−6,699	−22,192	−28,944				
Resi. SE (df = 7184)				0.110	0.107	0.150	0.147
F Stat (df = 15; 7184)				11,493.6***	10,415.0***	8,688.8***	8,936.0***

Note: †p < .1; *p < .05; **p < .01; ***p < .001
Standard errors of coefficients are displayed in the parentheses.

Table B.2: Robustness check on Experiment II (some suppliers learn faster) – simulation length 500

	Mixed-Effects					*OLS* (robustness check)	
DV: ln(AUC)	(R8) All	(R9) All	(R10) All	(R11) All	(R12) All	(R13) Honda	(R14) Rnd.1
Constant	5.589^{***}	6.001^{***}	6.110^{***}	6.064^{***}	5.942^{***}	6.090^{***}	6.130^{***}
	(0.026)	(0.026)	(0.028)	(0.011)	(0.037)	(0.012)	(0.015)
Diffusion (α)	0.783^{***}	0.253^{***}	0.084^{***}	0.146^{***}	0.300^{***}	0.107^{***}	0.056^{*}
	(0.005)	(0.008)	(0.024)	(0.011)	(0.022)	(0.023)	(0.027)
Disruption (β)	0.298^{***}	-0.011	-0.190^{***}	-0.122^{***}	-0.019	-0.182^{***}	-0.195^{***}
	(0.005)	(0.008)	(0.024)	(0.011)	(0.022)	(0.023)	(0.027)
Central	-0.051^{***}	-0.155^{***}	-0.193^{***}	-0.202^{***}	-0.343^{***}	-0.165^{***}	-0.201^{***}
	(0.002)	(0.009)	(0.018)	(0.008)	(0.017)	(0.017)	(0.021)
Baseline rate (a)	-1.902^{***}	-4.108^{***}	-6.600^{***}		-4.343^{***}	-5.912^{***}	-7.530^{***}
	(0.006)	(0.016)	(0.129)		(0.038)	(0.124)	(0.150)
Increased rate (a_x)	-0.054^{***}	-0.049^{***}	-0.162^{***}	$-0.017^{\dagger}$		-0.148^{***}	-0.186^{***}
	(0.003)	(0.009)	(0.018)	(0.009)		(0.018)	(0.021)
$\alpha \times a$		2.475^{***}	6.428^{***}		3.032^{***}	4.842^{***}	8.168^{***}
		(0.022)	(0.240)		(0.069)	(0.230)	(0.277)
$\alpha \times a_x$		0.040^{**}	0.207^{***}	0.011		0.141^{***}	0.279^{***}
		(0.012)	(0.034)	(0.017)		(0.033)	(0.039)
$\beta \times a$		1.894^{***}	6.062^{***}		2.425^{***}	5.404^{***}	6.959^{***}
		(0.022)	(0.240)		(0.069)	(0.230)	(0.277)
$\beta \times a_x$		$-0.022^{\dagger}$	0.157^{***}	-0.058^{***}		0.098^{**}	0.235^{***}
		(0.012)	(0.034)	(0.017)		(0.033)	(0.039)
$\alpha \times$ Centr		0.144^{***}	0.238^{***}	0.250^{***}	0.441^{***}	0.163^{***}	0.283^{***}
		(0.012)	(0.033)	(0.015)	(0.031)	(0.032)	(0.039)
$\beta \times$ Centr		0.089^{***}	0.188^{***}	0.192^{***}	0.225^{***}	0.191^{***}	0.167^{***}
		(0.012)	(0.033)	(0.015)	(0.031)	(0.032)	(0.039)
Centr $\times a$		0.190^{***}	-0.092		0.336^{***}	0.392^{*}	-0.539^{*}
		(0.023)	(0.183)		(0.053)	(0.175)	(0.212)
Centr $\times a_x$		-0.135^{***}	-0.158^{***}	-0.128^{***}		$-0.042^{\dagger}$	-0.267^{***}
		(0.013)	(0.026)	(0.013)		(0.025)	(0.030)
$\alpha \times \beta$			0.223^{***}	0.148^{***}	0.019	0.218^{***}	0.227^{***}
			(0.044)	(0.020)	(0.040)	(0.042)	(0.050)
$a \times a_x$			2.243^{***}			1.832^{***}	2.807^{***}
			(0.136)			(0.130)	(0.157)
$\alpha \times$ Centr $\times a$		-0.108^{***}	0.058		-0.380^{***}	-0.017	0.106
		(0.031)	(0.339)		(0.098)	(0.325)	(0.392)
$\alpha \times$ Centr $\times a_x$		0.190^{***}	0.212^{***}	0.173^{***}		$0.078^{\dagger}$	0.341^{***}
		(0.018)	(0.048)	(0.023)		(0.046)	(0.055)
$\beta \times$ Centr $\times a$		-0.184^{***}	-0.115		-0.461^{***}	$-0.548^{\dagger}$	0.125
		(0.031)	(0.339)		(0.098)	(0.325)	(0.392)
$\beta \times$ Centr $\times a_x$		0.024	0.041	0.027		$0.079^{\dagger}$	0.002
		(0.018)	(0.048)	(0.023)		(0.046)	(0.055)
$\alpha \times \beta \times$ Centr			-0.215^{***}	-0.218^{***}	-0.221^{***}	-0.168^{**}	-0.232^{**}
			(0.062)	(0.028)	(0.057)	(0.059)	(0.071)
Centr $\times a \times a_x$			0.443^{*}			-0.029	0.859^{***}
			(0.192)			(0.184)	(0.222)
...... Other higher-order terms omitted to save space. They are insignificant or not related to Central							
Observations	21,600	21,600	21,600	10,800	10,800	5,400	5,400
Adjusted R^2						0.992	0.993
Log Likelihood	12,039	26,108	27,283	22,196	11,811		
Akaike Inf. Crit.	$-24,063$	$-52,176$	$-54,498$	$-44,356$	$-23,586$		
Bayesian Inf. Crit.	$-23,999$	$-52,016$	$-54,227$	$-44,225$	$-23,455$		
F Stat (df = 31; 5368)						$23,026^{***}$	$23,346^{***}$

Note: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$
S.E. of coefficients displayed in parentheses.

B.4 *Post Hoc* Analysis on Centrality Metrics

Table B.3 presents the results of the *post hoc* analysis. All models (P1) through (P4) are mixed-effects models. Models (P1), (P2), and (P3) use numerical supplier learning rates in the regression, while model (P4) uses the categorized rates at three levels discussed in Section 4.6.3. We set the middle level $(a, a_x) = (0.2, 1)$ to be the reference in order to investigate a higher a and a lower a_x (easily comparison). Interaction terms are entered hierarchically. Model selection favors the full model with the highest log-likelihood and lowest AIC and BIC. Noticeably, both models (10) and (11) are full models with the only difference in the setup of supplier learning rates, hence returning equivalent log-likelihood, AIC, and BIC.

Out of the scope of the *post hoc* analysis, we investigate nodes with which centrality are critical to increasing rates on (i.e., from a to a_x) and under which condition. In theory, central suppliers are strategically more important with a higher number of paths through them. We examine the terms related to the out-degree centrality ("Out"; in-degree serves as the reference level).

Out-degree by its main effect helps improve supply network learning more than in-degree (-0.069^{***} in model (P1)), also visually verified in Figures 4.5 and B.3. It implies that fortifying risk spreaders (i.e., nodes with higher out-degree, which can affect others easily) in most cases is more important than helping vulnerable suppliers (i.e., nodes with higher in-degree, which are easily affected by others) – it is better to eliminate the source of the risks. The more important role of out-degree is determined in every situation. We then conclude that increasing learning rates for suppliers of higher out-degree always improves supply network resilience learning more, regardless of the factors. In other words, it is always better to try to eliminate risks from their sources.

Table B.3: *Post hoc* analysis results on centrality metrics

Mixed-Effects Model	Numerical learning rates			Categorical, ref: $(a, a_x) = (0.2, 1)$	
DV: ln(AUC)	(P1) All	(P2) All	(P3) All		(P4) All
Constant	5.643***	6.164***	5.911***	Constant	4.938***
	(0.060)	(0.062)	(0.065)		(0.062)
Diffusion (α)	1.444***	0.657***	1.164***	Diffusion (α)	1.881***
	(0.005)	(0.018)	(0.045)		(0.027)
Disruption (β)	0.802***	0.480***	0.986***	Disruption (β)	1.156***
	(0.005)	(0.018)	(0.045)		(0.027)
Out-degree	−0.069***	−0.073***	−0.053	Out-degree	0.050**
	(0.002)	(0.020)	(0.036)		(0.019)
Baseline rate (a)	−2.166***	−3.337***	−3.102***	$(a, a_x) = (0.2, 0.5)$	0.176***
	(0.009)	(0.035)	(0.064)		(0.018)
Increased rate (a_x)	−0.099***	−0.353***	−0.352***	$(a, a_x) = (0.5, 1)$	−0.931***
	(0.005)	(0.020)	(0.037)		(0.019)
$\alpha \times$ Out		0.078**	0.037	$\alpha \times$ Out	−0.275***
		(0.027)	(0.067)		(0.036)
$\beta \times$ Out		0.068*	0.027	$\beta \times$ Out	−0.076†
		(0.027)	(0.067)		(0.036)
$\alpha \times a$		1.453***	0.983***	$\alpha \times (0.2, 0.5)$	−0.260***
		(0.048)	(0.118)		(0.034)
$\alpha \times a_x$		0.522***	0.520***	$\alpha \times (0.5, 1)$	0.295***
		(0.028)	(0.068)		(0.036)
$\beta \times a$		0.779***	0.308**	$\beta \times (0.2, 0.5)$	−0.054
		(0.048)	(0.118)		(0.034)
$\beta \times a_x$		0.110***	0.108	$\beta \times (0.5, 1)$	0.092**
		(0.028)	(0.068)		(0.036)
Out $\times a$		−0.045	−0.106	Out $\times (0.2, 0.5)$	−0.062*
		(0.048)	(0.087)		(0.026)
Out $\times a_x$		0.108***	0.124*	Out $\times (0.5, 1)$	−0.032
		(0.028)	(0.051)		(0.026)
$\alpha \times \beta$			−1.012***	$\alpha \times \beta$	−0.821***
			(0.084)		(0.050)
$\alpha \times$ Out $\times a$		0.212**	0.334*	$\alpha \times$ Out $\times (0.2, 0.5)$	0.189***
		(0.065)	(0.161)		(0.047)
$\alpha \times$ Out $\times a_x$		−0.346***	−0.379***	$\alpha \times$ Out $\times (0.5, 1)$	0.100*
		(0.038)	(0.095)		(0.048)
$\beta \times$ Out $\times a$		0.060	0.182	$\beta \times$ Out $\times (0.2, 0.5)$	0.069
		(0.065)	(0.161)		(0.047)
$\beta \times$ Out $\times a_x$		−0.106**	−0.139	$\beta \times$ Out $\times (0.5, 1)$	0.054
		(0.038)	(0.095)		(0.048)
$\alpha \times \beta \times$ Out			0.082	$\alpha \times \beta \times$ Out	0.098
			(0.124)		(0.066)
$\alpha \times \beta \times a$			0.941***	$\alpha \times \beta \times (0.2, 0.5)$	−0.002
			(0.219)		(0.063)
$\alpha \times \beta \times a_x$			0.004	$\alpha \times \beta \times (0.5, 1)$	0.282***
			(0.126)		(0.066)
$\alpha \times \beta \times$ Out $\times a$			−0.244	$\alpha \times \beta \times$ Out $\times (0.2, 0.5)$	−0.032
			(0.299)		(0.088)
$\alpha \times \beta \times$ Out $\times a_x$			0.065	$\alpha \times \beta \times$ Out $\times (0.5, 1)$	−0.073
			(0.175)		(0.090)
Observations	10,800	10,800	10,800		10,800
Log Likelihood	9,068.5	11,715.5	12,468.3		12,468.3
Akaike Inf. Crit.	−18,121	−23,391	−24,885		−24,885
Bayesian Inf. Crit.	−18,063	−23,245	−24,695		−24,695

Note: †p < .1; *p < .05; **p < .01; ***p < .001

Standard errors of coefficients are displayed in the parentheses.

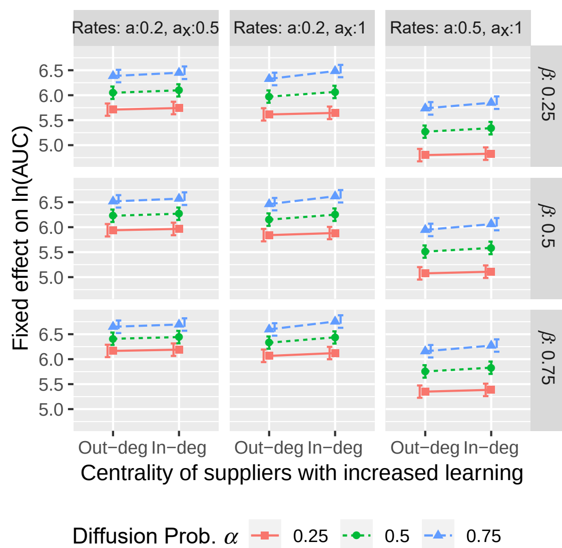


Figure B.3: Interaction effects between centrality, risk, and supplier learning in *post hoc* analysis