USING SIMULATED ANNEALING TO IMPROVE THE INFORMATION DISSEMINATION NETWORK STRUCTURE OF A FOREIGN ANIMAL DISEASE OUTBREAK RESPONSE

by

JAMES D. PLEUSS

B.S., United States Military Academy, 2007

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Industrial and Manufacturing Systems Engineering

College of Engineering

KANSAS STATE UNIVERSITY

Manhattan, KS

2016

Approved by:

Major Professor
Dr. Jessica L. Heier Stamm

Copyright

Abstract

Communication is an integral part of emergency response, and improving the information dissemination network for crisis communication can save time, resources, and lives. This thesis focuses specifically on emergency response to a foreign animal disease (FAD) outbreak, an incident in which an animal disease that is not active domestically is introduced and being spreading in the U.S. In a FAD outbreak, timeliness of detection and response are critical. An outbreak of foot-and-mouth disease, a particularly significant FAD, could cripple the agriculture economy and every hour of poor communication could result in the loss of thousands of animals. Improving this and other such crisis communication networks is of high importance. There is a comparatively large amount of prior research that critiques past catastrophic events but very little that aims to quantitatively improve such networks.

This research uses communication data from a FAD response exercise in Kansas to develop a reliable network model, contributing a general method for creating an information dissemination network from empirical communication data. The thesis then introduces a simulated annealing heuristic to alter the network structure, reducing the overall information transmission time by almost 90%. Both the application of simulated annealing in network design and the use of discrete event simulation to calculate the heuristic objective function are new contributions to the field of crisis communication and emergency response.

This work begins by extracting data from communication logs, grouping the large numbers of stakeholders into more manageable clusters, and developing a simulation model framework that accurately depicts the flow of information in the actual network. Then a simulated annealing heuristic is used to alter the network structure. The goal is to identify an alternative network structure in which the time for information to reach all response participants is minimized. The resultant network structures are analyzed to reveal observations and recommendations for FAD response communication.

This research finds that not only can such a network be improved significantly, but the quantitative results support the qualitative observations from early in the data extraction process. This paper adds original methods to the literature and opens the door for future quantitative work in the area of crisis communication and emergency response.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First of all, I would like to thank Dr. Jessica Heier Stamm for her support, encouragement, and incredible insight throughout this research. Her ability to understand what I am thinking without my ability to properly articulate it—and the patience required therein—is beyond impressive. I would also like to thank Dr. Todd Easton, who advised me on my initial program of study, and Dr. Jason Ellis for their comments and direction as part of my committee. Stephanie Coombes also helped immensely in the extraction of the data for this project. I would also like to recognize Sandy Johnson from the Kansas Department of Agriculture for providing the data and motivation for this research. Finally, thanks to LTC Randall Hickman for encouraging me to work with Dr. Heier Stamm in my research and helping me choose a program of study that made my time in the IMSE Department very enjoyable.

Dedication

This work is dedicated to my incredible and gorgeous wife, Ally, who supported me in every way throughout my time here at K-State. Her willingness to put up with my school, thesis writing, and never-ending list of house projects while raising our two beautiful daughters is something that I cannot understate. I love you, babe!

1 INTRODUCTION

In a world full of uncertainty and chaos, crises abound. From natural disasters to epidemics to acts of violence and terror, there is no shortage of events that require emergency coordination and operations. In any catastrophic event, getting the right information to the right people in an expeditious manner is paramount to the proper execution of emergency response. Foreign animal disease (FAD) outbreak responses constitute a type of event in which multi-stakeholder communication is important, and this context is the focus of this thesis.

Motivated by the need to better understand and improve crisis response communication, this thesis introduces a network model that represents the dissemination of information among stakeholders. Nodes in the network represent individuals or organizations, and arcs between them represent the flow of information. Given a model of an information dissemination network, the thesis then introduces a local search heuristic to decrease the time required to get information to responding organizations. The methods are demonstrated using data from a simulated FAD outbreak of foot-and-mouth disease (FMD) in the state of Kansas in October 2013.

This thesis begins with a discussion of the FAD exercise at hand as well as the goals and contribution of this research in Chapter 1. Chapter 2 summarizes related literature specific to crisis communication and information dissemination networks. Chapter 3 introduces a framework to generate a network simulation model using text-based data from communication logs, emails, call logs, web postings, and other communication data. Chapter 4 then expands upon this simulation using simulated annealing to develop an improved set of network structures. After a discussion of the numerical results of these structures in Chapter 5, the thesis concludes in Chapter 6 with final observations, recommendations, and areas for future research.

1.1 Background

The motivation for this thesis comes from three areas of significance: foot-and-mouth disease (FMD) and its catastrophic nature as a FAD, information dissemination network modeling, and the Kansas FAD response exercise in 2013.

1.1.1 Foreign Animal Disease

A FAD is defined by the United States Department of Agriculture (USDA) Animal and Plant Health Inspection Service (APHIS) as a disease affecting animals that is not currently active in the United States

[2]. The diseases that threaten animal or human health the most are divided into three tiers. Tier 1 FADs are the most catastrophic in nature, having the potential to impact the entire nation from one domestic case. FMD is considered a Tier 1 FAD.

FMD is an extremely contagious viral disease. It affects animals with divided hooves (cows, pigs, sheep, goats, deer, etc.) but is not a threat to public health or food safety. The illness manifests itself in blisters around the mouth, mammary glands, and hooves. While the disease itself does not often result in death, the infected animals are too weakened to produce meat or milk as before [3].

The FMD virus lives in the saliva, excrement, urine, and even breath of infected animals. The virus can also survive in contaminated materials, like bedding, for several months. This means that the virus can spread through any number of activities including transportation in contaminated vehicles, drinking contaminated water, contacting people with contaminated clothing/equipment, and being held in contaminated facilities. The survivability and numerous transmission modes show why this disease is so feared among agricultural stakeholders [3].

FMD was eradicated in the United States in 1929 and an outbreak has not happened since [4]. However, other countries with high exports of susceptible animals have had recent outbreaks. Taiwan had an outbreak of FMD in 1997 after its eradication in 1930, resulting in an economic loss of $1.6 billion and the slaughter of 40% of the at-risk population of pigs [5]. Then, in 2001, the United Kingdom had an outbreak resulting in 7 million animals slaughtered and $12–$18 billion in losses. Estimates for the total economic impact of a U.S. outbreak range from $23–$34 billion, caused primarily by cuts in both domestic and international demand [1].

1.1.2   Response Time and Information Dissemination

This research focuses on the flow of information among the key players and stakeholders in a FAD outbreak response. A FAD outbreak response is broken down into five elements [6]:

1. <u>Incident Identification</u>. Incident identification involves the detection of the outbreak symptoms, typically by the animal owner or producer and confirmation of the disease by lab results.

2. <u>Incident Management</u>. Incident management describes the establishment of emergency operations agencies and organizations to include Emergency Operations Centers (EOCs) at the state and local levels and determining a chain of command for the event.

3. Communication and Coordination. Communication and coordination includes establishing lines of communication with all involved stakeholders across the governmental, industry, and private sectors.

4. Assessment, Control, and Containment. This includes monitoring the spread of the disease, determining what areas/herds require quarantines, and restricting movement of susceptible animals as needed.

5. Recovery. Recovery occurs after the disease is eradicated and involves hazard evaluations, restoring movement, and reviewing the actions taken for future improvement.

The expediency of detection, effective and efficient response, and information dissemination play a large role in the size and scope of an outbreak. Carpenter *et. al* [4] examine the effect of delayed detection time on outbreak size and economic impact for a FMD outbreak in California. They found that as detection time increased from seven to 22 days, the number of animals slaughtered went from 8,700 to 260,400, and the economic impact to California rose from $2.3 billion to $69.0 billion. They also found that if a 21-day detection is assumed, every additional hour of delay would result in 2,000 animals slaughtered and $565 million in economic loss. However, the nature of the industry and variation in human contact with the animals makes improving detection time difficult, thus the time from infection to detection is highly variable [7].

Given the difficulty of rapid detection, effective response upon detection is essential. Response, including information dissemination, can be much more closely monitored and improved through practice and procedural changes. Kao *et. al* [5] discovered that, in the case of the United Kingdom FMD outbreak in 2001, if a nationwide movement ban had been imposed one day after discovery instead of three, then the number of infected populations could have decreased by more than 50%. This indecision was independent of the time taken for detection yet had a high consequence. Information dissemination is critical to the decision makers' ability to make timely and informed choices.

Part of this critical information flow lies with producers reporting potential infections to their veterinarians and following important orders such as stopping movement of cattle. Delgado *et. al* [8, 9] reveal results from two surveys showing these tendencies of producers. In general, they found that producers feel a lot of social pressure to report cattle with FMD symptoms and do understand the positive consequences of early reporting, but their understanding of the symptoms is lacking. Producers also are likely to follow stop movement orders during an outbreak due to social pressure and the fear of causing further spreading. Orders to gather and hold cattle, which helps for analysis and depopulation efforts, are less likely to be followed because of a lack of knowledge of the benefits and lack of resources to accomplish the tasks.

1.1.3 Kansas FAD Response Exercise

If this was not sufficient motivation for the state of Kansas to prepare for such an event, Lee *et. al* [1] look at the state-by-state impact of an FMD outbreak in California and its effects on the rest of the nation. Figure 1 shows the five states with the greatest potential economic loss (in millions of dollars). The lower and upper bounds refer to the lowest and highest losses over the four scenarios in [1]. Note that while Kansas has the lowest agriculture gross domestic product (GDP) of the five states, its percentage loss is significantly higher than the rest of the states. Kansas is estimated to lose 40–60% of its agriculture GDP in the event of a FMD outbreak. In fact, its lower bound is higher than the upper bound of all states except Nebraska. Making matters worse, only seven states (to include Nebraska and Iowa) have a higher percentage of their state economy coming from agriculture. This shows the significant impact of an FMD outbreak on the state of Kansas and why conducting exercises to improve communication and enforce standard operating procedures is important.
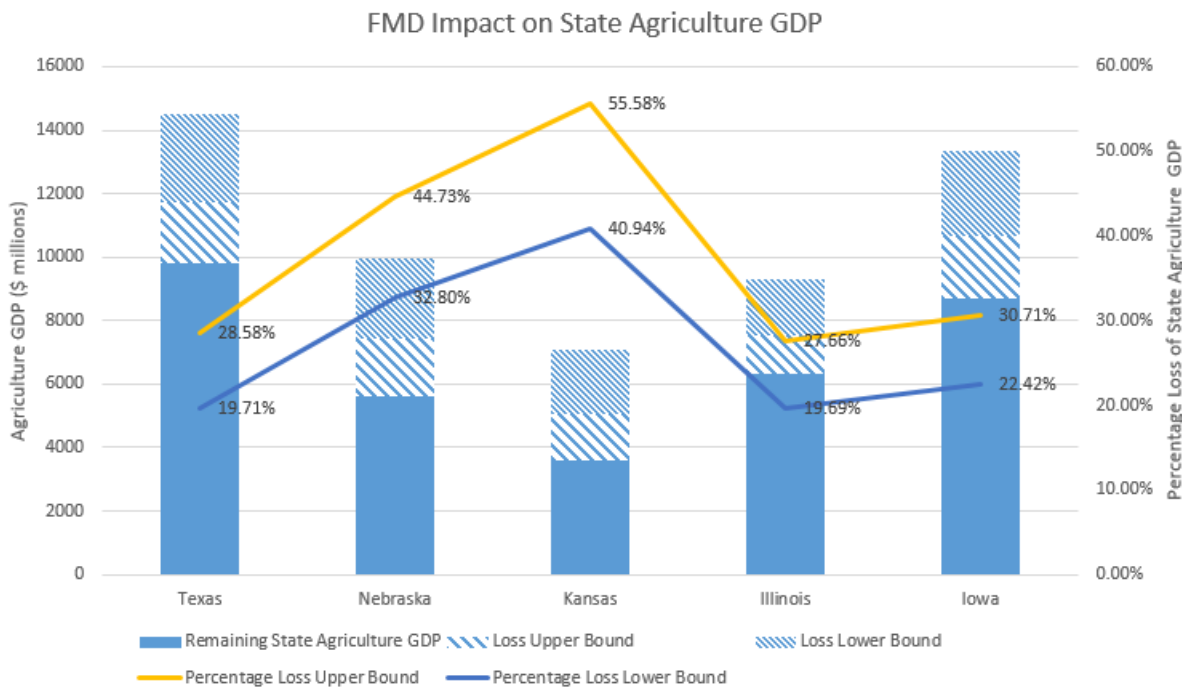
Figure 1: Comparison of the Agricultural Impact on the Five States with Highest Economic Loss from FMD Outbreak. Created from Data Presented in [1]
.

In October 2013, the Kansas Department of Agriculture (KDA) conducted a FAD response exercise engaging key stakeholders who would be involved in the event of an actual FAD emergency. This particular

exercise involved a simulated outbreak of FMD in the United States and the movement of an infected animal into the state of Kansas. The exercise began at 8:00 AM on October 9, 2013, and ended the next day at 1:00 PM with an overnight pause in the exercise from 5:00 PM to 8:00 AM. The stakeholders, or players, in the exercise represented many organizations and individuals on the federal, state, local, and industry levels. Local organizations included county emergency management personnel, law enforcement, and administration for the most impacted counties in the exercise scenario including Pottawatomie and Riley counties. At the state level, KDA was represented along with the state veterinarian, the Kansas Department of Health and Environment (KDHE), the Movement Control Branch (MCB), the Kansas Department of Emergency Management (KDEM), and state veterinarians from other nearby states. The federal level was represented by the Animal and Plant Health Inspection Service (APHIS), a division of the United States Department of Agriculture (USDA). Kansas State University also played a significant role in the exercise due to simulated events surrounding the College of Veterinary Medicine (CVM) and the Animal Sciences and Industry (ASI) Department. During an emergency event, many of these organizations and key players co-locate to central EOCs. In this exercise, there were multiple EOCs in play. The state had its own EOC, called the Kansas State Emergency Operations Center (KSEOC), each county had an EOC, Kansas State University had one, and, within KSU, the Animal Science and Industry Department had its own EOC.

An exercise like this is sensitive in nature because publicizing it could lead to public misperceptions or feelings of insecurity. For this reason and the availability of certain players or organizations, the exercise does not incorporate everyone that would otherwise have a role in an actual outbreak. An example of this is at the industry level where the various industry agencies are represented by an industry Multi-Agency Coordination Center (MACC), located with KDA. Typically, this MACC will have members of many different organizations such as the Kansas Farm Bureau (KFB), Kansas Livestock Association (KLA), and Dairy Farmers of America (DFA), to name a few. In this exercise, a member of KDA represented the MACC in the KSEOC. In short, those stakeholders with a role in a FAD response event but not participating in the exercise were represented by personnel within KDA.

The notional background of the exercise is that a farm in Alabama had a Wagyu bull present the symptoms of FMD after selling a heifer and calf from this herd to a farm in Kansas. The heifer was injured in transit to Kansas and went to the KSU CVM for evaluation where she, too, presented symptoms of FMD. The rest of the exercise is a response to this situation. Because some important entities did not directly participate in the exercise, injects for key events were used to represent actions taken in the event. These injects serve to input information into the exercise environment to observe how the players react to this new information.

For instance, the inject that begins the exercise is APHIS notifying KDA of a presumptive positive FMD in Alabama. Other injects occur throughout the exercise in order to observe and evaluate the responses of the players.

The observers and moderators of this exercise were assigned by KDA to the most important locations and personnel in the exercise. Their purpose was to evaluate how well these personnel perform their duties associated with a FAD response and document the communications throughout the exercise. Some of the locations with a dedicated observer were KDA, KSU Animal Science and Industry EOC, KSEOC, and MCB. The county EOC communication was logged through a web-based discussion portal.

1.2 Research Goals

This research is motivated by the need for KDA and similar agencies to better understand and improve crisis response. The author hypothesizes that the flow of information can be greatly improved through practical alterations in the communication patterns of those involved in a FAD response exercise or actual event. The first goal of this paper is to propose a generalizable modeling framework based on empirical data to better understand and manage crisis communication networks that support response efforts to foreign animal disease. The approach consists of a network model and discrete event simulation. Then, after developing an accurate simulation model for this response network, the second goal is to identify a new network structure that improves the information flow in the crisis communication network. This is accomplished with a simulated annealing heuristic (a type of local search) in which the objective function is calculated using the discrete event simulation model.

1.3 Research Contributions

This research relates to other areas in the literature, including past research on emergency response communication, information dissemination networks, and local search heuristics. It makes the following contributions:

1. *First research to be quantitative and prescriptive in analyzing official interactions in crisis communication.* This research fills an important gap in the literature of crisis communication. To the author's knowledge, it is also the first publicly available research analyzing a FAD emergency response communication network of any kind.

2. **Provides a method for creating an information dissemination network model from empirical communication data of varying qualities.** There is a shortage of literature on crisis communication using raw data and only a few instances known to this author where an actual communication network model is developed from such data. Within these, none create a model with the purpose of simulating that network. This marks the first such model based on empirical data.

3. **Uses simulated annealing to alter a communication network.** While simulated annealing is a very research-rich topic in local search heuristics, its application to network design is not. In fact, the only work related to network alteration with simulated annealing found by this author was in the field of reverse logistic network engineering, supply network design, and water supply networks [10, 11, 12]; none was found in the field of communication networks.

4. **Uses simulation to calculate the objective function value within simulated annealing for the purposes of network design.** To the author's knowledge, only one other piece of research uses simulation in determining the simulated annealing objective function [12]; that work does not use it in the process of complete network construction, but to add an additional element to an existing network.

5. **Offers insight into the communication network for a FAD emergency response in Kansas.** This research is the only effort to construct models based on this exercise communication data and provides insights specific to the FAD response network that are not addressed elsewhere in the literature.

These elements are presented throughout this thesis specifically in dealing with the Kansas FAD response of 2013, but the principles herein can be generalized and used in many applications of information dissemination networks.

2 LITERATURE REVIEW

The research contributions of this thesis lie in the intersection between multiple streams of literature that each address components of crisis response communication networks. What makes this thesis unique is that it simultaneously integrates four characteristics not previously combined in the literature.

1. **Communication.** It relates specifically to communication and information flow versus other types of network models.

2. **Quantitative.** The proposed methods are quantitative in nature rather than qualitative.

3. **Prescriptive.** It is prescriptive instead of descriptive. This means that this research aims to give a method or solution to be used in the future and not solely describe attributes of past events.

4. **Official.** The interactions it studies are primarily among official responders, rather than the general public or social networks. The data are drawn from official communications and not taken from social media or other unofficial sources.

There are five areas of research to be reviewed in this chapter, each of which exhibit a subset of these four core characteristics. Figure 2 shows how these areas reveal a gap in the research, which this thesis aims to fill. This chapter describes this gap, examining each of the five research areas in relation to this paper. In addition to these five areas, the chapter begins with network definitions and ends with a summary of the simulated annealing methodology.

2.1 Network Modeling Characteristics

Critical to this research is an understanding of network modeling and its many capabilities. A network is a group or system of interconnected entities (people or things). A network model is a structured depiction of the network using graphs that consist of nodes and arcs/edges/links. Nodes are entities within the network and usually depict people, places, organizations, and the like. Link is a generic term for the interaction between nodes that can refer to both edges and arcs. Arcs are one-way interactions between nodes, while edges refer to bidirectional connections. Thus, a network is modeled on a graph $G = (N, A)$, where N is the set of nodes and A is the set of all arcs (or E for the set of edges). A network model is a good way to approach this research because FAD response efforts involve individuals and organizations that can be represented as nodes while their communication with each other are the arcs. This modeling framework facilitates quantitative analysis using existing methods [13, 14, 15].
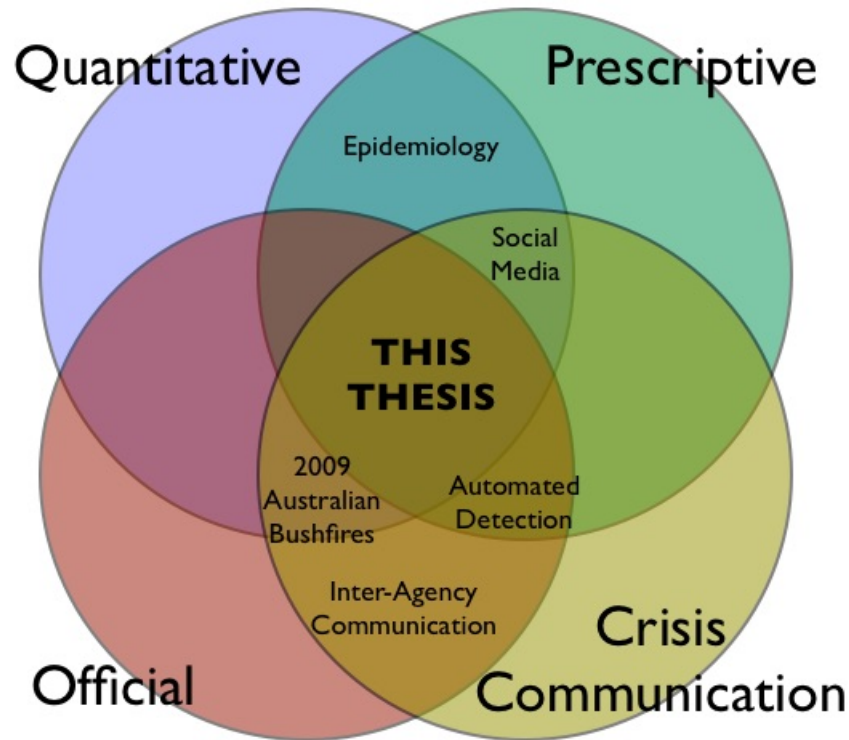
Figure 2: Gaps in the Literature Filled by This Thesis
.

Numerous metrics and characteristics are proposed in the literature to describe network structures. Among them are various measures of centrality, or how connected a node is with the rest of the network [16, 17]. Two other concepts are brokers and bridges. This paper considers the role of brokers and bridges when analyzing the network structures in the KDA response case study.

The difference between brokers and bridges is subtle but important; brokers are nodes or actors that provide interaction between two clusters of nodes, while bridges are the edges or communication links that span these gaps [18]. In cases where knowledge is transmitted across a network, a knowledge broker provides the critical link to bring useful information from one group to another. Having a broker to relay information comes at a cost, however; congestion at this broker can result in inefficient information flow [19].

With this general understanding of network models and some of their characteristics, along with how this research aims to use them, the focus now shifts to the five research areas in the literature most closely associated with this thesis in order of increasing relevance.

2.2 Automated Detection and Technology

Automated detection and technology, as it relates to crisis communication in the literature, is prescriptive and generally denotes official interactions, but it is not quantitative. It ties in with this thesis because one of the goals of this research is to improve the flow of official communication in an emergency response network. The prescriptive characteristic (improved response for future emergency events) is joined with crisis communication in the literature in the field of automation and technology, where organizations experience improved detection and response times because of technological advances in areas such as oil refineries and nuclear plants [20, 21, 22, 23]. This paper does not consider direct applications of technology for improvement of the network, but the literature suggests that there are trends toward exploring what benefits technology can bring to crisis communication efforts.

2.3 Social Media

Past research in the area of social media network models is quantitative, prescriptive, and can apply to crisis communication, but the interactions that are studied are not official. Areas within social media that relate particularly to this research are the ability to turn communication instances into quantifiable data and methods to evaluate how information moves through the network. As a definition, social media refers to "forms of electronic communication through which users create online communities to share information, ideas, personal messages, and other content." A social network is "a network of social interactions or personal relationships [24]."

As it relates to the four characteristics of this thesis in the beginning of the chapter, social media is used in crisis communication analysis often. A search of scholarly articles using the keywords "Social Media" and "Disaster" returned more than 600 articles and 30 reviews on the topic. Some of these articles are also prescriptive and discuss how social media can be used to detect future disasters like wildfires [25]. Research also shows benefits to using social media for official governmental agencies in crisis response to determine the needs and gravity of the event [26]. While terrorists can use social media to initiate and exacerbate an attack, public health officials can use social media to improve early warning and protect the people [27].

Finally, the closest relationship of social media to this thesis is when the literature moves from qualitative discussion to quantitative analysis. Sjoberg *et al.* [26] look at the use of social media in such bioterrorism incidents, finding a need for a multidisciplinary approach across law enforcement and health communities on the implementation of social media data mining. By analyzing keywords in Twitter messages, these

organizations can gain invaluable information before, during, and after a crisis. For comparison, this method is quantitative, used officially, and still in the realm of crisis communication, but it is descriptive instead of prescriptive.

Another quantitative relationship involving social networks and disasters comes from analysis of cell phone usage during emergency events. The method of extracting phone records to determine a network structure is very similar to the method used in this thesis. Gao *et al.* [28] use mobile phone data to quantify the communication patterns of users during a variety of anomalous events—a bombing, jet scare, concert, and plane crash. The study does not focus on those directly affected, but on the individuals the affected group contacts during the event. Do contacts call back to those affected to offer comfort or do they forward the information on to others and how do these actions compare to normal day activity between the two groups? The authors conclude that communication with the eyewitness is more critical (by virtue of representing a larger proportion of the communication) than forwarding on situational awareness for emergency events (the concert is not an emergency event).

Overall, social media plays an increasingly significant role in disaster communication and has been studied in kind. Due to its growth in the literature and its fundamental purpose of communication, research in social media has some models that relate quite closely to this research. Some of these principles can be applied to the more official communication in the network model associated with this paper.

2.4   Inter-agency Communication

Generally speaking, research related to inter-agency communication lacks quantitative analysis and prescriptive recommendations, but the interactions that are studied are official in nature and relate directly to crisis communication. A FAD response is dependent on varying organizations working together, communicating, and making good decisions in concert with one another, making inter-agency communication a close comparison despite not covering two of the four core characteristics. Even so, the literature has significance in that it reveals the importance of the work in this thesis. Specifically, it emphasizes the importance of sharing knowledge between organizations and discusses the disproportionally low number of quantitative models of these processes.

One of the most well known and documented catastrophes exacerbated by poor communication and information flow was Hurricane Katrina and its aftermath. Garnett and Kouzmin [29] discuss these communication deficiencies at length while also discussing the four communication lenses in crisis communication: interpersonal influence, media relations, inter-organizational networking, and technology showcase. Provid-

ing further motivation for our crisis communication study, they go on to point out that of these four lenses, the inter-organizational perspective is the least studied in terms of crisis communication and has great potential for improvement given this research gap. Another event with well-documented communication findings is the 2012 Tongariro eruptions in New Zealand. Leonard *et al.* [30] conducted a case study of the information flow within this event and focused on the interdisciplinary aspect of assessment and communication regarding the eruption. The study concluded that there is significant value in collaborative, multi-disciplinary environments with pre-existing relationships because they allow for efficient and unified public communication, whereby limiting public confusion and distrust.

A pair of reviews of public health emergency preparedness literature, [31, 32] provide confirmation that, despite a recent increase in research, the need for empirical and quantitative analysis in public health preparedness communication is high. More than half of the publications are reviews as opposed to primary research and less than 20% are quantitative in nature. In the case of inter-organizational communication modeling—where research is very limited—this percentage is even smaller.

Perhaps the most pertinent of articles in this section, [33] offers a good merger of the concept of inter-agency communication with the actual modeling of its characteristics. It relates to this thesis by comparing multiple strategies in its simulation model, just as this thesis does in the implementation of its simulated annealing heuristic. [33] differs from other works in this section because it is both prescriptive and quantitative. However, it does not relate directly to crisis communication but focuses instead on inter-organizational partnerships. Their model is based on empirical data of a known inter-organization collaboration network. They use agent-based modeling to simulate organizational decisions by looking at the missions and values of 119 organizations and 30 known collaboration projects. Their method looks at four different organizational strategies and four unique networks to develop the best possible collaboration network (thus, being prescriptive). Similarly, the FAD response network is improved in this thesis using simulation across many possible networks with four strategies to determine their structures.

Inter-organizational communication is a critical research area with room for quantitative models to complement existing qualitative methods. This last example shows how a quantitative approach can result in prescriptive results.

2.5   Epidemiology

Epidemiology modeling is the one area of research considered here that is completely outside the field of crisis communication because its quantitative and prescriptive characteristics can be utilized to also model

information dissemination, a key part of this research. The spread of disease can be viewed similarly to the spread of information [34]. While many items (for example, physical goods) move through a network in such a way that there are no duplicates, information and diseases both propagate by making a copy of the information/disease being passed before sending it along. This represents the principle that you do not lose information by passing it along. Likewise, the disease does not leave the host when it is passed on to someone else.

The backbone of such epidemiological models are their contact networks, just as the backbone of this thesis is the network created by communication data. Disease spread is often modeled using stochastic processes where individuals alternate between a discrete number of states. The most common depiction of this is the Susceptible-Infected-Susceptible (SIS) model, where individuals begin in a susceptible state to an infection and move to the infected state according to a probability distribution, only to return again after the infection passes [35, 36]. Similarly, this thesis models information dissemination where nodes (people, organizations, etc.) transition from being uninformed to informed. A variation of the SIS model is the Susceptible-Alert-Infected-Susceptible (SAIS) model, which adds in an alert state, wherein the infection rate changes due to behavior changes [35]. The alert state is determined by proximity to infected or alert individuals in the contact network among the susceptible and infected populations. This alert state brings epidemiological modeling even closer to this current research by adding in the alert (informed) state.

The same authors then take their research from being descriptive of the epidemiological network to prescriptive [37, 38]. In [37], they use heuristics, similar to this thesis, to search through potential contact adaptation patterns to find the one that minimizes the arcs needed in the information dissemination network while still effectively mitigating the epidemic. Finally, their most recent research [38] produces an optimal information dissemination strategy that results in greater suppression of the disease over a wider range of infection rates using fewer resources. This research combines these two ideas by looking for optimal information dissemination efficacy through the alteration of the contact network.

2.6   2009 Australian Bushfires

The quantitative analysis of official interactions in crisis communication during past emergency response events is the closest research field to this thesis, albeit a small one. In fact, after searching multiple scholarly databases for studies with these three core characteristics (quantitative, official, and crisis communication), the only relevant research is related to the Australian bushfires in 2009. The empirical communication data from this event among official responders sets it apart from other such events, and relates it more closely

to this thesis than other research. Empirical data for this type of event is needed to justifiably model a communication network because human decision-making is very difficult to predict [39], especially when faced with crisis circumstances that can easily cloud or shift an otherwise logical choice. This thesis builds on [40] and [41] by not only developing the network with empirical data, but adding a prescriptive characteristic in determining a better solution for the future.

Au [40] shows how to use this empirical data to develop a network model and then analyzes this network using two different network metrics, both important principles of this thesis. The nature of bushfires provides a repository of time-stamped communications between responders that is translated to a network model. Au organizes the nodes by organization (like the clusters developed for the FAD response in Section 3.2.1) and develops a complex network where the weight of the arcs represents the number of communication instances between the nodes (again, just as in 3.2.2). He then uses sociometric status and centrality as network metrics with which to analyze the network's performance. Sociometric status measures node activity based on the number of nodes it contacts (out-degree) and the number of times it is contacted (in-degree) relative to the overall communication in the network. These measures reveal some important gaps in the communication network at various phases of the fires, and while this thesis does not use these metrics specifically, the principle of using a metric to quantitatively evaluate an entire network applies.

Hamra *et al.* [41] take the bushfire data and generalize it in a way that is even more applicable to this thesis. They apply social network analysis (SNA) to determine whether the network interactions of actors help drive the success or failure of bushfire responses. They use communication data from two separate bushfire responses, one of which was far more successful than the other, to evaluate which aspects of the network (density, different types of centrality, and specific interactions) most impact the effectiveness of the response. This thesis also looks at the impact specific interactions have on the network's performance in Section 5.2.3.

With all five research areas reviewed, the most significant articles and their relationships to the four characteristics of this thesis are shown in Table 1. None of these articles satisfy all four characteristics, further motivating the need for this research.

2.7 Simulated Annealing and Network Design

This thesis uses a local search metaheuristic, simulated annealing, to identify an improved network for the FAD response exercise. A local search heuristic is a method of solving hard optimization problems by using a neighborhood structure to search for the best solution [42]. At any iteration, it considers a small well-

14

Table 1: How the Most Relevant Articles Relate to the Four Characteristics of This Thesis

Article Citation	Crisis Communication	Quantitative	Prescriptive	Official
[27]	X	X		X
[28]	X	X	X	
[33]		X	X	X
[37]		X	X	
[40]	X	X		X
[41]	X	X		X

defined change from the current solution to get a new solution. The neighborhood concept means it develops a list of feasible "neighboring" solutions from which one can be chosen and evaluated. A metaheuristic is a heuristic, or improving process, that has a method for moving away from a local optimal result [42]. There are many heuristics and metaheuristics in the literature for finding locally optimal solutions. Pezzini *et al* [43] explain the process for the most commonly used metaeheuristics including genetic algorithm, ant colony, particle swarm, tabu search, and simulated annealing.

The simulated annealing heuristic comes from Kirkpatrick's seminal work that draws on an analogy to the process of annealing metal, where metal is heated to allow for changes in its atomic structure and then slowly cooled in order to create the desired crystalline structure. As the metal goes from liquid to solid, the atoms lose their fluidity and that part of the structure cannot be changed [44]. Similarly, the heuristic allows for acceptance of a worse solution with decreasing probability as the heuristic progresses.

There is little research using simulated annealing to develop optimal network structures, and none relating to information dissemination or crisis communication. [10] uses simulated annealing to develop a better supply chain network in reverse logistics, a growing field in logistics. Chibeles-Martin *et al* [11] use simulated annealing to construct a locally optimal supply chain but do not use simulation modeling to determine their objective functions. In contrast, Samora *et al* [12] use simulation modeling for their simulated annealing heuristic objective function in finding the optimal locations for turbines in a city water supply network, but they do not design the entire network itself, only specific locations for new items. This research is the first to combine these two strategies to develop a simulation-based simulated annealing heuristic that determines a locally optimal network design structure.

2.8   Conclusion

A review of the scholarly literature in crisis communication and emergency response clearly indicates there are many opportunities for further research and development in this discipline. The limited research

is evidence for how challenging quantitative analysis can be for such a dynamic and subjective concept. Nonetheless, the benefits of conducting this research are great and many of these authors voice the need for this advancement.

3 NETWORK SIMULATION MODEL AND RESULTS

In light of the research already done in the field of crisis communication as well as the gaps recognized therein, this research aims to provide an effective method for determining an improved information dissemination network structure for emergency management operations using empirical data. Specifically, this research shows how to use the many types of communication logs gathered from an animal disease outbreak response exercise to effectively model its information dissemination network. This structure can then be modified to create a better conduit for information flow. The methods introduced in this thesis are generalizable to crisis communication networks at large.

This chapter begins by describing the framework of the proposed model through the definition of important network principles and an example information dissemination network used in the case study. It then describes the empirical data and the process developed to extract necessary information from these records, translating communication instances into a network structure. The model details then are described, after which the chapter presents results and validation. The final product is a communication network simulation based on empirical data that observes information from multiple sources as it flows throughout the network. The terms "model" and "simulation" are used inter-changeably in this paper because the model of the problem is done by means of a simulation.

3.1 Network Definition

This section introduces definitions and notation associated with the network model of information dissemination during a FAD response. It also provides the foundation for the primary metric used in this thesis.

3.1.1 Graph Components

Information dissemination among FAD response stakeholders is modeled on a graph $G = (N, A)$, where nodes in N represent participants (individuals or agencies) and arcs in A represent communications between participants. Arc (i, j) models information transmitted from node i to node j. The weight w_{ij} associated with arc (i, j) represents the strength of the connection between nodes i and j; pairs of nodes with higher weights on the arc connecting them are expected to communicate more frequently than those with lower weights. This information is summarized in an n x n weighted adjacency matrix W, where $n = |N|$ and w_{ij} is the element in the ith row and the jth column of W. Thus, for every position in W where $w_{ij} > 0$, there is

an arc in the network from i to j. Two arcs between a pair of nodes represent a bi-directional relationship. Similarly, an arc from a node to itself reveals intra-node communication. Finally, each node has total weight $w_i = \sum_{j=1}^{n} w_{ij}$.

There are often cases, such as the exercise this thesis addresses, where the nodes in the network represent more than a single person or player. In such cases, nodes representing many players may transmit information quicker than those representing only one individual. Additionally, if a node represents multiple players, it is possible that only a portion of them have received the information moving through the network. Modeling these intricacies of the network is important because they could be a very effectual part of the data set. In order to capture these facets of the network, each node, i, is assigned a maximum capacity λ_i^{max} and current capacity λ_i. λ_i^{max} is assigned as the total number of players represented by node i. The current capacity is initialized as 0 for each of the nodes and increased each time a node receives information. Figure 3 shows an example network.
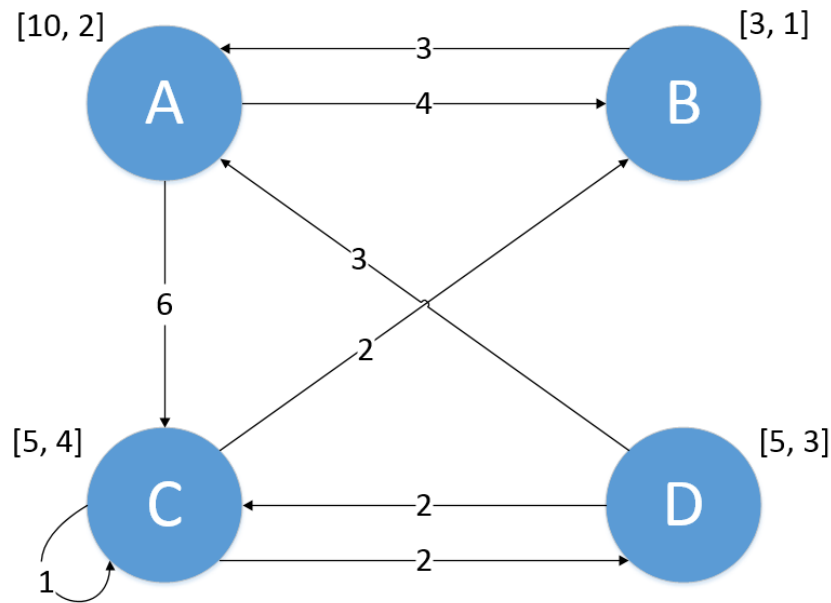
Figure 3: Diagram of Example Network; label on arc (i, j) is w_{ij} and label on node i is $[w_i, \lambda_i^{max}]$

3.1.2   Information Flow

Information dissemination through the previously described network is modeled using discrete event simulation. In this model, information is transmitted from one node (called the "transmitting" node or

"sending" node inter-changeably) to another "receiving" node. At the beginning of the simulation, one node is declared to be the originator of the information (more on how this is determined later in the chapter). The receiving node is determined using a probability matrix, P. In this matrix, $p_{ij} = \dfrac{w_{ij}}{w_i}$, where p_{ij} is the probability of transmitting node i choosing to transmit to node j.

The actual implementation of this method uses a related rate matrix, R, such that

$$
R = \begin{cases} r_{ij} = \dfrac{w_{ij}}{w_i}, & \text{for } j = 0 \\[2ex] r_{ij} = r_{i(j-1)} + \dfrac{w_{ij}}{w_i}, & \text{otherwise} \end{cases},
$$

and a uniformly distributed random number from 0 to 1 (u[0,1]), produced by a Random Number Generator (RNG). The cumulative nature R for each sending node, i, makes the determination of the receiving node, j, easier to find when using a u[0,1]. The model progresses through the transmitting node's row until it finds a value that is greater than or equal to the u[0,1]. That column's index, j, is the receiving node of the transmission. Tables 2 and 3 give the probability and rate matrix for the example network in Figure 3.

Transmission times for each node are determined according to network paramters, as described in detail in Section 3.3.3.

Table 2: Probability Matrix for Example Network

P Matrix	A	B	C	D
A	0	0.4	0.6	0
B	1	0	0	0
C	0	0.4	0.2	0.4
D	0.6	0	0.4	0

Table 3: Rate Matrix for Example Network

R Matrix	A	B	C	D
A	0	0.4	1	1
B	1	1	1	1
C	0	0.4	0.6	1
D	0.6	0.6	1	1

3.1.3 Informed

An important concept in this research is the definition of the term, "informed". The "inform(ed)" descriptor can be applied to either individual nodes or to entire networks alike. A node is considered informed when its maximum capacity has received the information passing through the network, or, mathematically, when $\lambda_i = \lambda_i^{max}$. Similarly, a network is informed when all nodes are informed, or $\lambda_i = \lambda_i^{max}$ for all $i \in N$.

Using the example network above, consider the following series of events. The information to be transmitted originates at node A. This means that $\lambda_A = 1$ at the start of the simulation while λ_B, λ_C, and λ_D are each 0. Because $\lambda_A^{max} = 2$, node A is not yet informed. Our RNG will provide a u[0,1] to determine the node to which A will transmit. Suppose the random number is 0.25. In this case, node A will transmit to

node B because r_{AB} is the first value in row A of the rate matrix, R, where $r_{Ai} > 0.25$. This makes $\lambda_B = 1$ and, because $\lambda_B^{max} = 1$, node B becomes informed. The simulation continues based on the next transmission times of the nodes until all nodes and, thus, the network, are informed. The time at which the network is informed is the primary metric in this thesis.

3.2 Converting Data to a Model

The components and parameters of the network simulation model are derived from crisis communication data. The data for this research come from documented communications during a simulated FAD incident in the state of Kansas. The Kansas Department of Agriculture serves as the focal point of the exercise. Its work includes assigning observers (also called moderators) to evaluate the actions of the important personnel in the FAD response. These moderators log many communication instances throughout the exercise and provide observations of successes and areas that need improvement. KDA also develops a schedule of injects for the exercise. An inject is the introduction of a new element to the environment, such as a simulated phone call from a citizen, and is a way for the observing organization to test how well the emergency response personnel deal with potential circumstances should such an event actually occur.

The communication records for this research come in the form of call logs, moderator comments, simulation injects, and copies of various emails, all in different levels of organization and clarity. The next step is to use this data to construct a network model that can be used to understand and improve the communication system.

3.2.1 Determining Clusters

In a FAD response, many entities are involved, spanning all levels of government and numerous civilian and industry-related sectors. In a large-scale FAD exercise like this, the players involved can be in the hundreds to thousands. In an actual event, when the public and news media become involved, that number is far greater. It is not reasonable in these scenarios to expect accurate or comprehensive communication logs for each player involved. Additionally, even with that sort of data, altering the structure of the network in a way that reveals general observations of the structure and information flow would be very challenging. One method to make a detailed network like this more generalized is to divide the players into groups, or clusters [13], which are then represented as nodes in the information dissemination network.

One approach to creating clusters is the use of a heuristic. Clustering heuristics use network data and are applied after the data extraction phase [45, 46, 47]. The KDA scenario does not have enough information

Table 4: Labels for the Network Nodes

Label	Node Name	Label	Node Name	Label	Node Name
A	County/Local Administration	**G**	Kansas State Emergency Operations Center (KSEOC)	**M**	Local Veterinarians
B	County/Local Emergency Management	**H**	Kansas State University (KSU)	**N**	Movement Control Branch (MCB)
C	County/Local Law Enforcement	**I**	Kansas Department of Agriculture (KDA)	**O**	News Media
D	Federal	**J**	Kansas Department of Health & Environment (KDHE)	**P**	Other States
E	Individual Producers	**K**	KSU Animal Science & Industry (KSU ASI)	**Q**	Public
F	Industry Associations	**L**	KSU College of Veterinary Medicine (KSU CVM)	**R**	Unknown

on individual players to create an accurate network depiction. In instances of small or sparse data, expert knowledge of the system is advantageous and can be applied to create clusters before or after data extraction.

Clusters for the KDA case study are generated prior to data extraction. Table 4 provides the resulting network nodes and abbreviated labels for easier referencing. Most of these nodes are self-explanatory but the Unknown (**R**) node may require further explanation. In the raw data of the communication logs, there are a few communication instances where the transmission does not have a distinguishable recipient. Posting to a Web Log is one example of this type of communication. There also are some transmissions where the recipient is known, but the source is not. Even without complete information, these transmissions are important and throwing them out would make the network model less representative of the actual communications. To account for them, they are assigned to an Unknown node with a maximum capacity of 1 so it does not slow down the network's informed time but still effectively models these transmissions.

3.2.2 Data Extraction Process

The data extraction process for this research is quite involved and takes a lot of time, depending on the quality of the data and communication logs. The order of operations for developing this network structure is specific to this FAD exercise but can be generalized and used for other information dissemination networks with similar data.

1. Compile all applicable data logs, communication logs, and any standard operating procedures or emergency management policies that help inform proper reporting procedures.

2. Use the existing documented policies and procedures to develop a list of possible organizations, individuals or positions within those organizations, and the clusters to which those organizations belong. Individuals and positions are both encapsulated by the term player in this thesis. The reasoning is that there are instances where a position is filled by multiple individuals but specifying each person is not necessary. For instance, an organization may have a role that rotates daily among multiple people, in which case the importance in the communication is that it comes from that position and not from the individual. There also are instances where an individual's name is not provided, in which case the position must suffice. This is often the case in standard operating procedures that aim to generalize the requirements to account for turnover of personnel.

3. Create a master communication log, called the Extracted Transmission List, from the collection of communication and data logs tracking the following: time of transmission; the organization, position/name, and cluster for both the sending and receiving players; and the mode of communication (email, phone call, conference call, etc). It also is helpful to include what was said in order to track how specific information flows through the network. Only the time of transmission and sending/receiving clusters and positions will be used in developing the network structure for this particular thesis. However, the rest of the extracted information could be used for future work or may be more helpful for other information dissemination scenarios.

   The roles of sending and receiving players are significant and warrant further discussion. For general purposes, it is assumed that the one making the phone call, sending the email, or initiating the conference call, is the one passing along information. However, there are instances where the initiator of the communication (e.g., the person who makes the phone call) is actually on the receiving end of the information flow. In these cases, the sending and receiving players should be logged according to the actual flow of information. An example of this from the exercise is the Public (**Q**) node. In all but one instance of the actual exercise, the Public initiated communication with another node, but it was asking for information, not transmitting new information. This is difficult to distinguish in most transmissions because the content of the conversation is not usually well-documented. But when it is clear, as in the case of the Public node, the transmission log should follow the flow of information.

4. Reorganize the Extracted Transmission List by time of transmission.

5. Set w_{ij} as the number of transmissions from node i to node j and λ_i^{max} as the number of unique players associated with node i in the Extracted Transmission List.

For a more detailed example of this process, see Appendix II. In all, there are 431 transmissions from more than 70 documents with the players in these transmissions broken down into 18 distinct clusters. Table 5 shows the resulting weight matrix, W, for the network, as well as the total weight for each node, w_i, and the maximum capacity for each of the nodes, λ_i^{max}.

Table 5: Weight Matrix and Capacities for the Network

W	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	w_i	λ_i^{max}
A	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	4
B	1	4	0	0	0	0	2	2	6	0	0	0	0	0	0	0	5	20	40	20
C	0	0	7	0	0	0	1	0	1	0	0	0	0	30	0	0	0	0	39	16
D	0	3	0	0	0	0	1	1	16	0	0	1	0	0	1	10	0	0	33	13
E	0	0	0	0	0	0	0	0	9	0	0	1	1	0	0	0	0	0	11	6
F	0	2	0	0	0	3	1	1	14	0	0	0	0	0	0	0	0	0	21	14
G	0	1	0	0	0	0	2	0	11	0	0	0	0	0	0	0	0	6	20	11
H	0	3	0	0	0	0	0	6	3	0	5	2	0	0	2	0	7	0	28	9
I	0	13	0	0	0	11	11	3	14	2	7	9	0	1	3	8	20	1	103	13
J	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	8	0	9	1
K	0	0	0	0	0	0	0	4	5	0	26	0	0	0	0	0	0	0	35	7
L	0	1	0	3	1	0	1	0	5	0	1	4	0	0	0	0	3	1	20	9
M	0	4	0	0	0	0	0	0	3	0	0	2	0	0	0	0	0	0	9	5
N	0	0	8	0	0	0	0	0	7	0	0	0	0	6	0	5	0	0	26	10
O	0	2	0	0	0	1	0	0	5	0	0	2	0	1	0	0	2	0	13	9
P	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	9	8
Q	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	30
R	0	3	0	0	0	0	0	3	2	0	0	0	0	0	0	0	0	0	8	1

Additionally, Figure 4 gives a visual representation of this network with the thickness of the arcs corresponding to their weight.

3.2.3 Observations of KDA Exercise Data

The final representation of the KDA exercise data in Table 5 and Figure 4 brings up some initial observations of the network to consider:

- KDA is the biggest hub by far in the network, both in receiving and sending information, yet it does not have a comparatively large capacity to some of the other nodes. This could indicate that it is not structured to be able to sustain its communication demands.

- Federal receives all of its information from the KSU CVM but sends information to many nodes.

- The Movement Control Branch and County/Local Law Enforcement are very closely related.
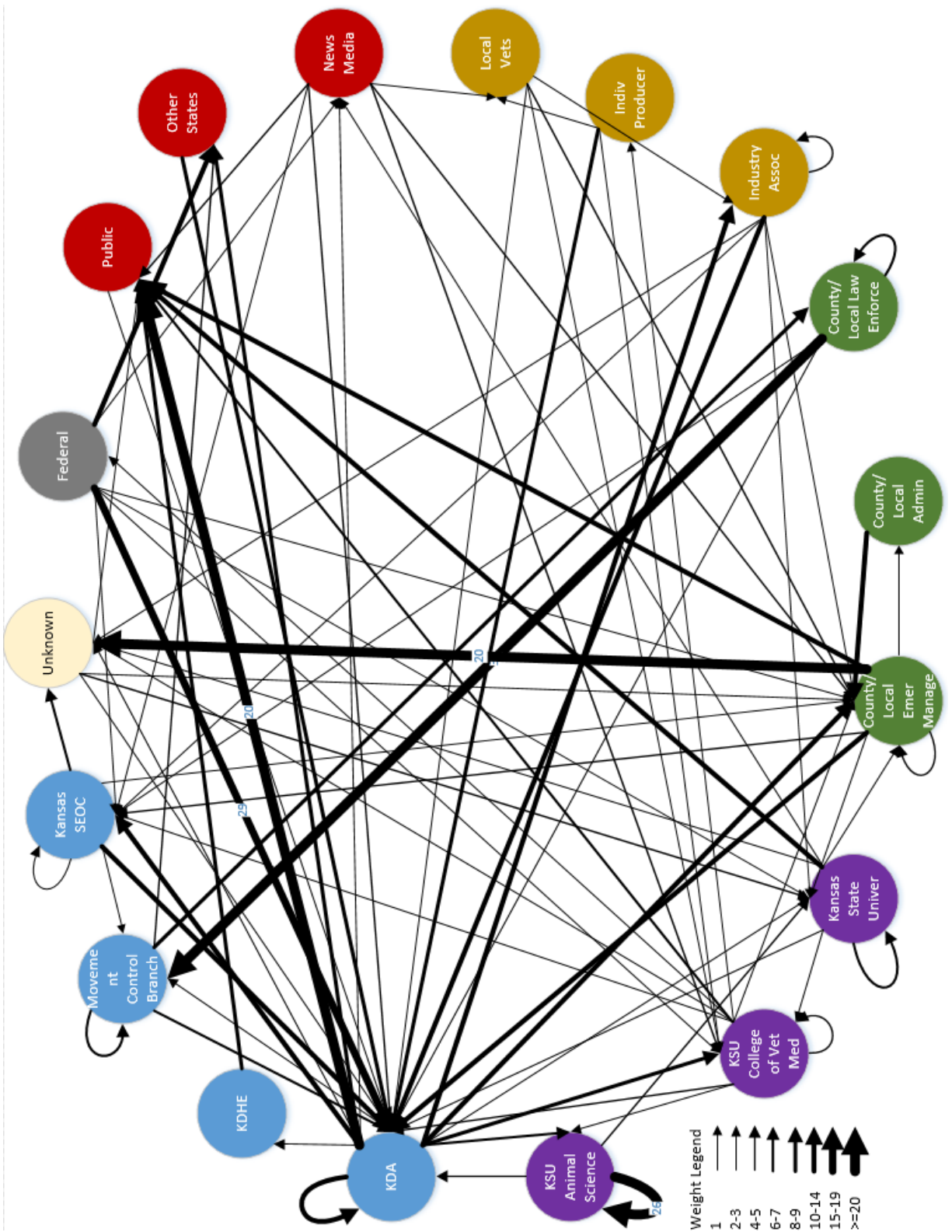
Figure 4: Graph of Network. Blue Nodes are State Nodes, Green are Local, Yellow are Industry-Related, and Purple are KSU Nodes.

- KDHE, County/Local Administration, and Individual Producers all receive information from a singular source (KDA, County/Local Emergency Management, and KSU CVM, respectively).

- More intra-county communication is expected. This could be a byproduct of the large amount of County/Local Emergency Management to Unknown links. This could be a result of the exercise design that had limited counties participating.

- Industry Association unexpectedly has no communication with the Individual Producers.

3.3 Simulation Design and Parameter Selection

Given the network representation of the FAD response data, the next objective is to simulate the flow of information through the network. This section describes the selection of nodes from which information originates in the simulation, the sequence of simulation events, and the processes used to select simulation parameters.

3.3.1 Information Injects

The simulation model depicts the flow of information through three injects, each replicated numerous times. These three injects constitute pieces of information originating at three different nodes and represent the most common and important sources of information from the exercise itself. The first inject simulates information originating at KDA ($\mathbf{I}$), the coordinating organization of the FAD exercise. KDA is by far the most active of all nodes with more than twice the total weight in $w_{\mathbf{I}}$ than any other node. It also introduces the most new information to the FAD exercise of all nodes, making it a natural choice to originate information in the model. The second inject choice represents initial detection, the most critical piece of information in an FAD outbreak. In most cases, the symptoms of a possible FAD, like foot-and-mouth disease, are originally noticed by the owner of the animal, represented in this network as an Individual Producer ($\mathbf{E}$). The critical nature of this piece of information is why this node is chosen as one of the injects. Finally, another common source of information in the exercise is the Federal node ($\mathbf{D}$), which passes along requirements and directives to the various states impacted by the FAD. It also can represent information from an event in another region where entities are not otherwise connected to the nodes represented in this model. These three nodes are representative of the most common and critical sources of information in the FAD exercise. It would be feasible to model information origination from all network nodes or even just one, and some applications

of these methods may warrant such decisions, but for this particular exercise, the three selected injects are most relevant.

3.3.2 Sequence of Simulation Events

Algorithm 1 shows the pseudo-code for this simulation. In broad terms, the simulation begins at time 0 initializing the current capacity array for the nodes (set to 0), and the next transmission time for each node (set to infinity until the node receives the information for the first time). The current capacity of the originating node is set to one and its first transmission time is determined. The transmitting node is determined by the smallest transmission time of all the nodes. At the transmission time of a node, the transmission recipient is determined using the u[0,1] number and the transmitting node's associated row in the R matrix. The receiving node's current capacity is incremented by one (unless it is already at λ^{max}), and the transmitting node receives a new next transmission time. The receiving node will also receive a next transmission time if its current time is still at infinity. The simulation loops through the transmission times until all nodes have reached their maximum capacity, making the entire network informed.

Algorithm 1 Simulation Pseudo-Code

1: Define number of replications as $replications$ and $rep = 0$
2: **for** inj as **D**, **E**, and **I** (to represent the three inject origination nodes) **do**
3:     **while** $rep < replications$ **do**
4:        $time = 0$. Number of informed nodes, $informed = 0$.
5:        Initialize current capacity, $\lambda_i = 0$ and transmission time, $T_i = \infty$ for all $i \in N$.
6:        Increment capacity of origination node $\lambda_{inj} = 1$.
7:        Get transmission time for origination node T_{inj}.
8:        **while** $informed < 18$ (number of total nodes) **do**
9:           Find earliest transmission time T_i.
10:           $time = T_i$.
11:           Determine recipient j.
12:           **if** $\lambda_j < \lambda_j^{max}$ **then**
13:              Increment λ_j.
14:              **if** $\lambda_j = \lambda_j^{max}$ **then**
15:                 Increment $informed$.
16:                 **return** informed time for node j.
17:              **end if**
18:           **end if**
19:           Get new transmission times T_i and T_j.
20:        **end while**
21:        **return** network informed time for replication rep and inject inj
22:     **end while**
23: **end for**

3.3.3　Determining Next Transmission Time

The most mathematically intensive part of the simulation is determining the next transmission time for each of the nodes. The next transmission time is assumed to be a random variable with a distribution estimated from the FAD exercise data. This section describes the process used to select probability distributions for each node.

The data used to find a distribution for the model are the inter-transmission times. Inter-transmission times are determined for a node i by finding all the differences in time between subsequent transmissions where node i is the transmitting node. Each node i has $w_i - 1$ inter-transmission times.

1. A theoretical distribution (i.e., Normal, Exponential, Gamma, Weibull, etc.) should be used in finding the transmission times. Law [48] recommends the use of a theoretical distribution for simulations that require randomness. If a distribution does not fit the compared data, then an empirical distribution can be used. An empirical distribution is using the data itself to run the model, rather than drawing from a distribution that fits the data. The small number of observed transmission from certain nodes makes using an empirical distribution impractical, because nodes where the total weight is only 1 would always have the same transmission time.

2. Some combination of the frequency of a node's transmissions in the exercise and its capacity should influence its transmission time in the simulation. If all nodes use the same distribution and same parameter(s) then they perform identically, which would not be representative of the system itself. Likewise, for the modified networks discussed in Chapters 4 and 5, it is necessary to choose distributions that depend on parameters that will be available for each node in each candidate network structure. Node i's capacity, λ_i^{max}, and total weight, w_i, are the only values available for each of the nodes.

3. Many nodes have too few observed transmission times in the exercise data to fit a distribution. Thus, the five nodes with at least 30 observations are used to select one distribution type. Then each node's parameters (λ_i^{max} and w_i) are used to specify the distribution of the selected type that is used for that node.

4. Goodness-of-fit (GoF) tests are used to determine how well or poorly a distribution may represent a given data set. Three such tests are typically used in this field: the Chi-Squared test, the Kolmogorov-Smirnov (K-S) test, and the Anderson-Darling (A-D) test. Each of these have their own strengths and weaknesses. In small data sets such as this one, the Chi-Squared test loses its validity [48]. Also,

for the K-S and A-D tests, the small sets limit the number of distributions that can be tested to the Normal, Lognormal, Exponential, Weibull, and Log-logistic distributions. Both of these tests have their benefits and can be used in this model. The K-S test looks at the greatest disparity in the data set from the distribution while the A-D test expands on this to give a higher importance to the tails of the given distribution function.

With these guidelines in mind, finding a distribution fit for the nodes with the highest number of inter-transmission times is the next step. These five nodes are County/Local Emergency Management (**B**), County/Local Law Enforcement (**C**), Federal (**D**), KDA (**I**), and Kansas State University Animal Science & Industry Department (**K**).

EasyFit software [49] is used to automate the goodness-of-fit testing. The null hypothesis is that inter-transmission times fit the distribution being tested. For a given data set and acceptable Type I error rate α, a test statistic is computed for each distribution. This is compared to a critical value. If the critical value is greater than the test statistic, then one fails to reject the null hypothesis at the level of α. Otherwise the null hypothesis is rejected at the same α level. Failing to reject a null hypothesis is not the same as accepting it as true. This is because a data set will never completely and perfectly fit a given distribution. The goal is to be reasonably close. For this research, the desire is to find a distribution that fails to reject the null hypothesis with an $\alpha > 0.10$.

Table 6: K-S Test Statistics Against Critical Values ($\alpha = 0.1$)

Node/Distribution	Critical Value	Exponential	Log-Logistic	Lognormal	Normal	Weibull
Node (**B**)	0.202	0.258	0.2099	0.2105	0.216	0.162
Node (**C**)	0.191	0.634	0.3788	0.43	0.393	0.683
Node (**D**)	0.208	0.644	0.356	0.427	0.337	0.341
Node (**I**)	0.134	0.466	0.299	0.334	0.276	0.301
Node (**K**)	0.202	0.686	0.424	0.427	0.413	0.388

Table 7: A-D Test Statistics With Critical Value 1.929 ($\alpha = 0.1$)

Node/Distribution	Exponential	Log-Logistic	Lognormal	Normal	Weibull
Node (**B**)	3.94	2.026	2.137	2.23	1.175
Node (**C**)	27.085	9.688	6.825	10.239	5.904
Node (**D**)	35.732	5.668	7.042	5.868	5.429
Node (**I**)	28.921	8.611	9.011	8.412	6.894
Node (**K**)	44.511	8.731	6.72	8.765	6.101

Tables 6 and 7 show the test statistics and critical values for the K-S and A-D tests, respectively, for each of the five nodes. No distribution sufficiently fits all of the nodes. In fact, node **B** and the Weibull

28

distribution is the only test for which the null hypothesis is not rejected. It appears at first glance that the Normal distribution may be a good choice because it consistently has lower K-S test statistics than the other distributions. However, the histograms of the data given in Figure 5 reveal characteristics that goodness-of-fit statistics do not show. Figure 6 illustrates the shapes of the five possible distributions.
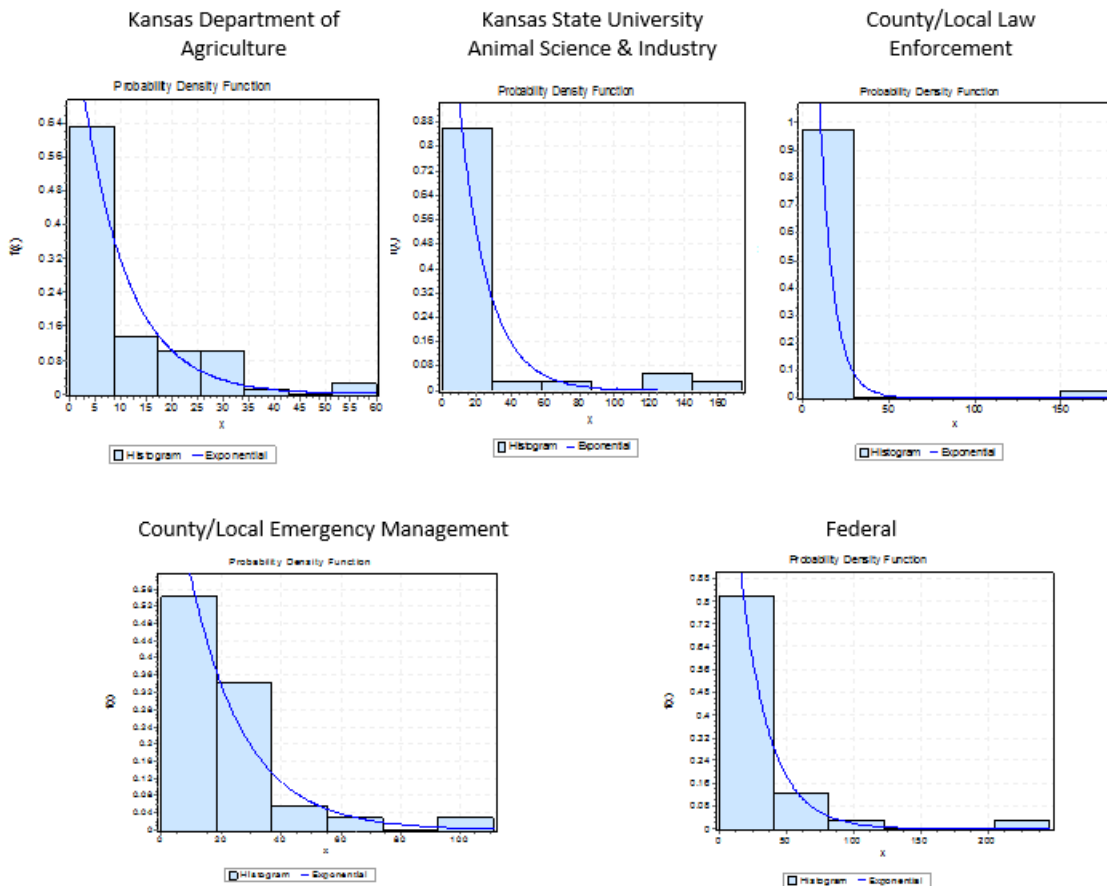
Figure 5: Transmission Times for Nodes with the Largest Number of Outgoing Transmissions

Comparing Figures 5 and 6 makes the exponential distribution look like the best choice of the five possible distributions. The blue curve in each of the graphs in Figure 5 is the exponential distribution fitted to that data. With insufficient statistical evidence to support a different distribution, and too few data points to use an empirical distribution, this thesis adopts the exponential distribution to model the transmission times.

3.3.4   Distribution Parameter

Given the choice of exponential distribution to model transmission times, this section describes how the distribution parameter is chosen for each node in the network model. The exponential distribution is
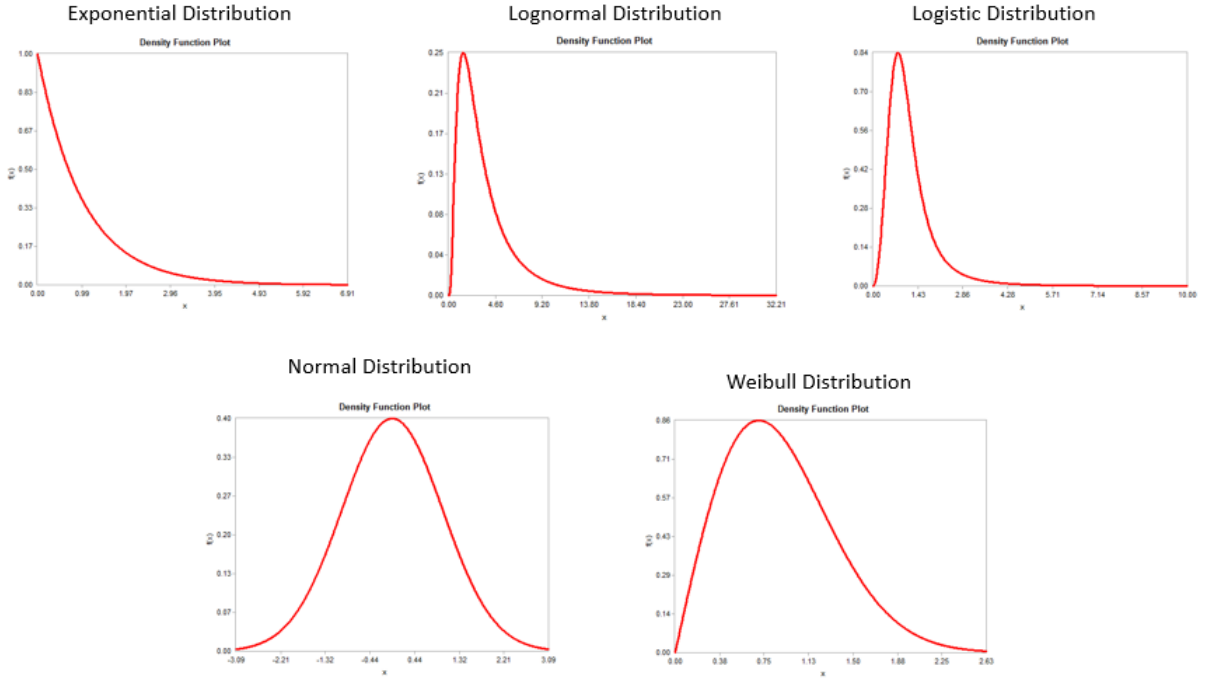
Figure 6: Density Function Plot of Allowable Distributions for K-S GoF with Low Sample Size

specified by a single parameter, denoted β_i for node i. The expected value of an exponential distribution is $1/\beta_i$ which means that as β_i increases, the expected time until the next transmission from node i decreases. This is important in considering how to determine the exponential parameter of each node as a function of known node parameters, weight w_i and maximum capacity λ_i^{max}.

The goal is to find a single function of w_i and λ_i^{max} that provides a reasonable β_i for every node. The results for the five nodes with largest weight are used to select this function. Table 8 summarizes the known node parameters, w_i and λ_i^{max}, and the best-fitting exponential distribution parameter, β_i, for the five nodes.

Table 8: Parameters of the Exponential Function

Node	w_i	λ_i^{max}	β_i
B	40	20	0.055
C	39	16	0.1198
D	33	13	0.0458
I	103	13	0.108
K	35	7	0.058

Six functional forms are considered, each of which admits a scalar z; these functional forms are provided in the first column of Table 9. Linear regression is used to select the value of z that minimizes the sum of squared errors between the predicted $\beta_i' = f(z, w_i, \lambda_i^{max})$ and the best-fitting value β_i for each of the

six functional forms. The best z and the corresponding sum of squared errors for each functional form are summarized in Table 9. The function $\beta_i' = 0.00133(w_i + \lambda_i^{max})$ yields the minimum sum of squared errors, and thus, this function is used to determine the exponential distribution parameter for transmission times for every node in the network.

Table 9: Sum of Squared Errors for Exponential Parameter

Parameter Function	z	Sum of Squared Errors
$z(\lambda_i^{max} * w_i)$	0.0001	0.00544
$z(\lambda_i^{max}/w_i)$	0.182	0.01293
$z(w_i - \lambda_i^{max})$	0.0021	0.00743
$z(w_i + \lambda_i^{max})$	0.00133	0.00344
$z(w_i)$	0.0017	0.00425
$z(\lambda_i^{max})$	0.0053	0.00708

3.3.5 Number of Replications

The final simulation design component is selecting an appropriate number of replications to enable reporting results with a desired confidence level. The result of interest for this study is the informed time of the entire network. Each simulation replication produces an informed time. To ensure with 95% confidence that the true average informed time is within 100 minutes of the observed average informed time from the simulation replications, an initial sample of 50 replications for each inject is taken. The required number of replications, r, is determined by the following equation: $r = \left(\dfrac{t_{\alpha/2} * s}{B}\right)^2$, where r is the number of runs required, B is the desired precision level (in our case 100), s is the sample standard deviation, and $t_{\alpha/2}$ is the critical value of the student-t test for a Type I error of α (in our case, 0.05) [50]. Table 10 shows the number of replications needed for each inject to reach the desired confidence. KDA requires the largest number of all injects at 367.07. For uniformity, and even stronger confidence, 400 replications are used.

Table 10: Number of Replications Required for 95% CI $\pm$ 100 minutes

Inject Node	KDA (**I**)	Individual Producers (**E**)	Federal (**D**)
Sample Average Informed Time	3619.23	3463.97	3669.71
Sample Standard Deviation	948.0	846.8	895.74
$r = \left(\dfrac{t_{\alpha/2} * s}{B}\right)^2$	367.07	292.88	327.72

3.4   Data Limitations

The nature of the case study data imposes some limitations on this analysis. These are divided into two categories, namely limitations arising from the fact that the data originate from an exercise rather than an actual event and those that arise from the data format or content. Some of the data set limitations are:

- Exercise limitations:

  - There is no documentation of communication for players that did not actually participate in the exercise but do play a significant role in a real-life scenario.

  - There are a few players or groups of players that could be their own cluster, based on their importance in the standard operating procedures and other literature, but that are never on the receiving end of any communications.

- Data format/content limitations:

  - Not all communications are properly documented, including all face-to-face conversations.

  - Some moderators are better at logging communication transmissions than others, making for some areas with disproportionately more documented transmissions than others of similar involvement in the exercise.

  - Information that is posted to a message board or online tracking system does not have a designated recipient, making those messages difficult to incorporate into the model.

These limitations may impact model accuracy, specifically how quickly a network or node becomes informed. As the number of documented transmissions for a node increases, the transmission time of the node decreases. However, an increase in transmissions would likely mean an increase in maximum capacity as well, making it impossible to identify whether more communications would necessarily improve network performance.

Validation of this network is also challenging because of the data. Specifically, there are very few specific pieces of information that can be tracked through the Extracted Transmission List and none that are documented as reaching every node. This makes determining if the model accurately reflects the actual network much more difficult.

There are many ways to improve upon this data set in future exercises, some of which have already been implemented by KDA as a result of this research. First of all, streamlining communication documentation

by specifying a common email, electronic log, or data portal can ensure more communications are being documented. Secondly, increasing the number of moderators will allow for a broader range of documented communication over more nodes and training them properly will help them understand the importance of specifying transmission times, members of the conversation, and content discussed. Despite these challenges in the data set, the Extracted Transmission List is still large and significant enough to create a model that depicts the general structure and layout of the communication network.

3.5   Model Validation

This section describes model validation efforts based on comparisons between the simulation outputs and the FAD exercise data. After 400 replications of this network structure on each of the three injects, the average informed time of the network was more than 3,000 minutes, or 50 hours, for each inject. Considering the FAD exercise itself was only 30 hours, this would appear to indicate an inaccurate network model. After all, how could a single piece of information never make it to the entire network over the course of the exercise? This is a reasonable question to ask, but remember that an informed network is one in which the current capacity of every node equals its maximum capacity. Unfortunately, the lack of data means there are no instances in the exercise itself where the information reaches every player represented, making proper validation of the model difficult.

Ideally, validation for this system would include following a single piece of information from one of the inject sources through the Extracted Transmission List until it reaches every player at every node, noting the informed times along the way and comparing them to the simulated results. However, as already discussed, there is no such information thread available. In an effort to provide the best alternative, there are two pieces of critical information that can be tracked through the exercise.

The first piece of information coincides with the beginning of the exercise where the Federal node tells KDA of a presumptive positive FMD case in the state of Alabama. There are 37 transmissions in the Extracted Transmission List that discuss this information, eventually reaching nine different nodes. This event is an example of inject 3 of the model, where the Federal node originates information to the network. The lack of documented transmissions means the actual informed times of nodes cannot be determined, but the time at which a node first receives the information, or "first informed time"(when λ_i becomes > 0), can be compared for these first nine nodes.

Table 11 shows this comparison in detail. Comparing the exercise and simulation times of each row suggests that the simulation is, indeed, representing this information flow in the network closely, particularly for

33

the first seven nodes, where the times differ on average by less than 15 minutes. Additionally, the simulation predicts the actual order of the first three nodes perfectly and the nine exercise nodes are represented in the first 13 nodes of the simulation.

Table 11: First Informed Times for Presumptive Positive FMD and Inject 3 over 400 Replications

Number of Nodes First Informed	Node	Exercise Time (min)	Node	Simulation Time (min)
1	D	0.00	D	0.00
2	I	0.00	I	25.78
3	P	0.00	P	41.51
4	O	54.00	Q	47.88
5	G	60.00	B	53.85
6	N	60.00	R	67.01
7	F	60.00	G	69.32
8	K	110.00	F	82.48
9	H	150.00	L	82.66
10			H	87.66
11			K	92.59
12			O	142.32
13			N	333.06
14			J	379.19
15			C	403.47
16			A	553.48
17			E	609.69
18			M	1029.17

The only other instance from the exercise with enough transmission documentation to analyze is when a bull with possible FMD is discovered in the Kansas State University College of Veterinary Medicine (**L**). There are 47 such transmissions for this event, though it reaches only seven distinct nodes. This event most closely resembles inject 2 of the model where an individual producer (**E**) introduces information to the network. With this in mind, Table 12 shows the same comparison as the previous instance. Just as in the previous example, all seven exercise nodes are represented within the first 11 nodes of the simulation. Also, the average difference in times for the individual nodes is less than 25 minutes, indicating another reasonable validation.

Differences between this simulation model and the actual data set are expected, particularly because of the randomized and probabilistic nature of the simulation. And while the limited data prevents a perfect comparison for validation, the two instances described in this section provide evidence of its fidelity. For the purposes of this thesis, the model developed throughout this chapter performs sufficiently well and the remainder of the thesis will build on this network model.

Table 12: First Informed Times for Bull at KSU CVM and Inject 2 over 400 Replications

Number of Nodes First Informed	Node	Exercise Time (min)	Node	Simulation Time (min)
1	L	0.00	E	0.00
2	I	0.00	I	42.54
3	K	108.00	Q	69.13
4	G	108.00	B	81.59
5	B	117.00	G	89.30
6	R	117.00	L	91.76
7	H	139.00	R	93.09
8			F	102.61
9			K	111.48
10			P	111.87
11			H	118.61
12			O	195.80
13			D	265.44
14			J	390.85
15			N	395.21
16			M	424.68
17			C	462.76
18			A	561.43

3.6 Simulation Results

The information dissemination network simulation model is run 400 times for each of the three injects described previously. The informed time for each node and for the whole network are recorded for each replication. Table 13 shows the average informed time for the 400 replications for each node and inject in the simulation as well as the average informed time of the overall network. These results provide insight about the current network structure and help identify future directions for improvement.

Interestingly, information originating at the Individual Producer (node **E**) moves through the network quicker than that originating at nodes with both higher total weight and maximum capacity (**I** and **D**). This is due to the decrease in informed time to the Local Veterinarians (node **M**), which are only contacted in the network by the Individual Producers. It also indicates a potential need to adjust the network in way that gives more links to Local Veterinarians.

Another intriguing result is the large difference between the average network informed time and the highest node informed times for each inject. For instance, the network informed time for inject 1 is 3661.94 minutes but the highest node time is Local Veterinarians (**E**) at 2953.90 minutes, a difference of almost 700 minutes. This shows the large variance in the network towards the end of the simulation. There are four nodes for inject 1 that have an average informed time more than 2,000 minutes (County/Local Administration,

Table 13: Average Informed Time for All Injects over 400 Replications

Node Informed	Inject 1 (min) From I	Inject 2 (min) From E	Inject 3 (min) From D
County/Local Administration(A)	2042.14	2017.39	2058.28
County/Local Emergency Management(B)	381.57	402.16	376.68
County/Local Law Enforcement(C)	934.73	995.51	929.97
Federal(D)	2248.90	2307.53	2068.25
Individual Producers(E)	2953.98	2604.49	3069.49
Industry Associations(F)	567.34	609.06	602.91
Kansas State Emergency Operations Center(G)	373.34	413.97	382.61
Kansas State University(H)	318.60	353.37	317.43
Kansas Department of Agriculture(I)	102.13	133.75	106.88
Kansas Department of Health & Environment(J)	348.24	390.85	379.19
KSU Animal Science & Industry(K)	176.43	209.09	190.45
KSU College of Veterinary Medicine(L)	339.54	345.07	349.66
Local Veterinarians(M)	2822.64	2118.96	2730.95
Movement Control Branch(N)	516.71	575.36	511.24
News Media(O)	990.36	1023.82	972.2
Other States(P)	354.12	390.87	252.33
Public(Q)	456.28	496.92	472.29
Unknown(R)	55.38	93.08	67.01
Entire Network	3661.94	3291.67	3757.31

Federal, Individual Producer, Local Veterinarians), and these four tend to alternate which node is the last to be informed. These nodes also exhibit greater variability in average informed times (average standard deviation of 968 minutes) than the other nodes (average standard deviations of 148 minutes). Again, this indicates a need to get these four nodes more involved in the network with regards to incoming communication links.

In Section 3.2.2, some observations were noted based on the network flow chart from Figure 4. The simulation results confirm and give greater understanding to some of those initial observations.

- KDA was identified as a hub of information and these results confirm that, showing that it has the second lowest average informed time despite having one of the highest maximum capacities in the network.

- Federal is very much affected by only receiving information from KSU College of Veterinary Medicine as it is one of the last three informed nodes for all injects.

- The Movement Control Branch and County/Local Law Enforcement are definitely closely related. In fact, the difference in informed times between the two across the three injects are 418.02 minutes $(934.73 - 516.71)$, 420.15 minutes $(995.51 - 575.36)$, and 418.73 $(929.97 - 511.24)$ minutes, respectively. Considering the individual informed times vary among the three injects by more than 60 minutes, this small variation in their difference is very indicative of a strong correlation between the two.

These discoveries reveal the need for an improved network structure. Without some changes, particularly to the nodes with the highest informed times, expeditious transfer of critical information to all players in this FMD response will be impractical at best.

3.7   Conclusion

This chapter explains the basics of information dissemination network structures and shows how raw data from communication logs can be used to develop a model that accurately depicts the performance of the network. Specifically, this chapter concludes with a reliable network simulation of the Kansas FAD exercise. The next chapter will explore how this network can then be altered in an effort to reduce the average informed time of the network.

4   IDENTIFYING IMPROVED NETWORK STRUCTURES

   While the focus of Chapter 3 was on building a network model based on actual communication data and simulating the performance of the current network structure, this chapter introduces a method to identify alternative network structures with improved performance. Improvement of this network requires the ability to test other network structures and accurately compare them against the original network and each other.

   The challenge becomes, with such an enormous number of possible networks, how does one determine which to test and what metrics should be used to compare them? This chapter introduces a simulated annealing heuristic to identify improved network structures. In addition, it describes the simulation principles and experimental design utilized in the implementation.

4.1   Simulated Annealing Algorithm

   This section introduces the main components of the simulated annealing algorithm. It includes the neighborhood, the objective function, the cooling parameter, and the stopping conditions.

4.1.1   Neighborhood

   The primary operation in a simulated annealing algorithm is selecting a new solution that is in the "neighborhood" of the current solution. A solution network structure is defined by its weight matrix. This section describes feasible network structures and the definition of neighbors. The original weight matrix used in Chapter 3 is denoted W^0 to set it apart from all subsequent network structures. These subsequent structures are denoted W^k, where k is the iteration number.

Feasible Network Structures

   There are a few structural requirements each network must meet in this simulation. When adjusting the weight matrix, it is important to consider what the changes may indicate if they are to occur in real life. In some information dissemination networks, it may be feasible for everyone to be able to communicate with everyone, but that is not the case for this exercise. For instance, the Federal node would not feasibly talk to the Local Veterinarian node in actuality. To ensure network structure changes that make logical sense, a n x n boolean potential adjacency matrix, A, is constructed where $a_{ij} = 1$ when transmissions from node i to node j are allowable and $a_{ij} = 0$ if not. This matrix is constructed based on expert knowledge of the system and how the various players communicate. Table 14 is the original weight matrix, W^0, from Table 5,

Table 14: Weight Matrix with Possible Adjustment Cells for the Base Network

W	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	w_i	λ^{max}
A	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	4
B	1	4	0	0	0	0	2	2	6	0	0	0	0	0	0	0	5	20	40	20
C	0	0	7	0	0	0	1	0	1	0	0	0	0	30	0	0	0	0	39	16
D	0	3	0	0	0	0	1	1	16	0	0	1	0	0	1	10	0	0	33	13
E	0	0	0	0	0	0	0	0	9	0	0	1	1	0	0	0	0	0	11	6
F	0	2	0	0	0	3	1	1	14	0	0	0	0	0	0	0	0	0	21	14
G	0	1	0	0	0	0	2	0	11	0	0	0	0	0	0	0	0	6	20	11
H	0	3	0	0	0	0	0	6	3	0	5	2	0	0	2	0	7	0	28	9
I	0	13	0	0	0	11	11	3	14	2	7	9	0	1	3	8	20	1	103	13
J	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	8	0	9	1
K	0	0	0	0	0	0	0	4	5	0	26	0	0	0	0	0	0	0	35	7
L	0	1	0	3	1	0	1	0	5	0	1	4	0	0	0	0	3	1	20	9
M	0	4	0	0	0	0	0	0	3	0	0	2	0	0	0	0	0	0	9	5
N	0	0	8	0	0	0	0	0	7	0	0	0	0	6	0	5	0	0	26	10
O	0	2	0	0	0	1	0	0	5	0	0	2	0	1	0	0	2	0	13	9
P	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	9	8
Q	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	30
R	0	3	0	0	0	0	0	3	2	0	0	0	0	0	0	0	0	0	8	1

but it also includes information from this adjacency matrix, where cell w_{ij} is highlighted if $a_{ij} = 1$. In other words, only highlighted cells are allowed to have a positive weight in future network structures.

The next limitation on alternate network structures is a zero-sum rule for outgoing weights. That is, any outgoing link weight deducted from w_{ij} must be added to some w_{ik} with the same transmitting node, i. For example, if a link weight of 1 is taken away from $w_{3,4}$ then 1 must be added to $w_{3,j}$ where j is any other node such that $a_{3,j} = 1$. The purpose of this, practically speaking, is to force a node to redirect its communication, not add more or simply stop communicating. This is easily put into practice because it considers the potential limitations on how many transmissions one cluster of players can actually make. The network also cannot be altered in such a way that a node no longer receives any communication from other nodes. Mathematically speaking, this requirement is represented as $\sum_{i=1}^{n} w_{ij} > 0, i \neq j$ for all $j \in N$.

Similarly, disjoint groups of nodes are not allowed. It is possible for a network change to result in a group of nodes that only communicate with each other, preventing the information from ever reaching them or, if the information originates in the group, never leaving. One solution to determine this is to conduct a breadth first search on the network. Instead, this thesis handles it with a limit on the maximum run time of a replication at 7,000 minutes. This number is chosen because the highest total run time of any replication in the base network is 7,003 minutes and this maximum time prevents a potential loop from distorting average informed time statistics. These rules ensure every network analyzed is feasible and the simulation of that

network will terminate in a reasonable amount of time.

Neighborhood Definition

Each network structure has a set of feasible neighbors. These neighbors are feasible in that they subscribe to the aforementioned limitations. In this thesis, four neighborhood variations are considered, named Swap 1, Swap 3, Swap 3-1, and Swap 3-2-1. The four variations relate to how much weight is changed in each iteration of the simulated annealing. A neighbor of the network defined by weight matrix W^1 is a network whose weight matrix W^2 satisfies $w_{ij}^2 = w_{ij}^1 - swap$ and $w_{ik}^2 = w_{ik}^1 + swap$ for exactly one pair j and k. To prevent negative weights, if $swap > w_{ij}^1$ then $w_{ij}^2 = w_{ij}^1 - w_{ij}^1 = 0$ and $w_{ik}^2 = w_{ik}^1 + w_{ij}^1$. As an example, when $swap = 1$, $w_{ij}^2 = w_{ij}^1 - 1$ and $w_{ik}^2 = w_{ik}^1 + 1$ for exactly one pair j and k. That is, the only differences between W^1 and W^2 are two individual cells whose weights differ by one. A neighborhood is the set of all new network structures that satisfy the above criteria.

By substituting another variable, $swap$, the size of this alteration can be adjusted. The idea is that a larger swap will lead to more drastic changes in the network structure. The neighborhood variations are denoted with a letter, A through D, corresponding to Swap 1, Swap 3, Swap 3-1, and Swap 3-2-1, respectively, added in the superscript of the weight matrix. So the network structure in the fifth iteration of simulated annealing for Swap 3 would be represented by W^{5B}.

Two of the neighborhood variations (Swap 1 and Swap 3) are related only to the $swap$ variable, where $swap = 1$ and $swap = 3$. The other two variations are considered a progressive shift, where $swap$ decreases as the simulation runs. For Swap 3-1, $swap$ changes from three to one after 50 consecutive non-improving iterations. Swap 3-2-1 decrements $swap$ from three to two after 25 consecutive non-improving iterations and from two to one after another 25. These different neighborhoods change how the model converges to its best solution and each provides a different final network structure as well.

4.1.2 Objective Function

The objective function for simulated annealing is the average informed time of the network across all 1,200 replications. This is calculated within each iteration of the simulation. As discussed in 3.6, there are many different statistical outputs provided by this model. Naturally, the focus is on those metrics that provide information on the entire network rather than individual nodes.

In a crisis communication scenario where timeliness and accuracy of information is paramount, a fully informed network is the safest and most conservative parameter. There are certainly other goals that could

be desired and a similar approach can be adopted for these. For the purposes of this simulation, however, the informed time for the network is the preferred metric with which to compare various structures because it is the only way of ensuring that all players requiring the information have received it. Centrality, as discussed in Chapter 2, was also considered as a potential predictor of network performance but further experimentation (explained in detail in Appendix I) found no correlation with the informed time of the network.

Expanding on the discussion of network informed times, all injects are equally important for the purposes of this research. There is no indication from the exercise that weighing the injects differently would make the model more accurate. Because of this, the primary metric used for this simulation with which to compare all iterations is the average informed time of the network for all 1,200 replications (400 replications for all three injects). The average informed time of a given network W^k denoted $f(W^k)$.

The goal of this simulated annealing method is to identify a network structure with very good $f(W^*)$ in a reasonable computational time. Finding a globally optimal network is also not feasible for the purposes of this research. Nonetheless, a locally optimal network W^*—the best network for all tested network structures— will still provide great insight into network improvement for this exercise.

4.1.3 Cooling Parameter

As the simulated annealing process continues, it "cools", decreasing the probability of accepting a worse structure. This probability at iteration k, ρ_k, is determined using the equation $\rho_k = e^{(-\Delta/c^k)}$, where Δ is the difference between $f(W^k)$ and the objective function value of the most recently accepted network W^A and c is the cooling parameter.

The cooling parameter c must satisfy $c < 1$. This is important because it drives the probability of acceptance to 0 as k, the number of iterations, increases. If $c = 1$, then the probability will depend only on Δ, potentially choosing a much worse network structure far into the simulation when the heuristic should be "fine tuning" the results. On the opposite side, if $c > 1$ the probability approaches 1 as k increases, resulting in choosing the new structure no matter what the objective function value is.

With the goal of conducting thousands of iterations, it makes sense to allow the acceptance of a worse network structure, even at the 500th iteration. Parameter tuning is done to identify an appropriate c value for this system. After a test of 500 iterations where $c = 1$ (so as to prevent the probability of acceptance from approaching 0 or 1), the average Δ for non-improving iterations was 32.06 minutes. What cooling parameter, c, will give an acceptance probability of 0.1 ($\rho_k = 0.1$) at the 500th iteration ($k = 500$) using $\Delta = 32.06$?

41

Using the equation above, $\rho_k = e^{(-\Delta/c^k)}$, and solving for c, reveals $c = \sqrt[k]{\dfrac{-\Delta}{ln(\rho_k)}} = \sqrt[500]{\dfrac{-32.06}{ln(0.1)}} = 1.005$. However, this violates the restriction of $c < 1$. In order to get the desired probability with an acceptable parameter, the Δ must be decreased. This is accomplished by adding a scaling factor, S, to the original equation, creating $\rho_k = e^{(-\Delta/Sc^k)}$. A scaling factor of 100, for instance, changes the required parameter from 1.005 to 0.996, satisfying the parameter restriction.

This means, however, that there are now two parameters to define before the model is complete. After some preliminary testing, S and c are first limited to values from 50 to 500 and 0.9 to 0.9999, respectively; anything outside of these ranges performs far worse. Next, combinations of these are tested for 400 iterations, each with the resultant optimal average informed time shown in Table 15. The best times for each S are highlighted in green. A cooling parameter of 0.99 results in the three lowest optimal average informed times in the matrix, with $S = 100$ having the best result at 671.62 minutes. Setting $c = 0.99$ and running for 1,200 iterations confirmed $S = 100$ as the top choice as shown in Table 16. Then, the cooling parameter is taken to further specificity in Table 17 for 15,000 iterations to find $c = 0.9995$. In the end, the parameter choices after these tests are $S = 100$ and $c = 0.9995$.

Table 15: Optimal Average Informed Time for 400 Iterations

Optimal Average Informed Time (min)	Scaling Factor (S)					
	50	100	200	300	400	500
0.9	820.09	809.68	764.39	859.01	730.16	744.39
0.91	851.46	832.35	820.51	931.16	794.11	936.00
0.92	851.46	873.24	804.30	931.16	781.09	894.68
0.93	851.46	873.01	799.74	930.63	783.44	881.96
0.94	851.46	836.86	794.12	771.20	792.69	731.30
0.95	828.08	813.32	830.56	796.04	791.79	817.51
0.96	763.34	806.10	821.45	1106.54	923.23	719.03
0.97	759.46	761.78	810.88	796.11	782.83	715.45
0.98	899.07	762.72	738.17	816.51	732.32	692.95
0.99	826.90	671.62	679.19	685.89	756.02	815.56
0.99999	808.04	763.84	835.21	875.51	873.11	1038.74

(Cooling Parameter (c) is the row label axis.)

4.1.4 Stopping Conditions

One aspect of the simulated annealing heuristic not yet discussed is when to conclude the network structure alterations. Common stopping conditions in the literature are a combination of maximum number of alternate configurations m and the number of consecutive configurations in which there is no improvement

Table 16: Informed Time at $c = 0.99$	
1,200 Iterations	
Scaling Factor, S	Time (min)
1	559.05
50	645.86
100	538.58
200	577.98
300	559.59
400	547.30
500	554.28
10,000	550.16

Table 17: Informed Time at $S = 100$	
15,000 Iterations	
Cooling Parameter (c)	Time (min)
0.992	463.55
0.994	458.64
0.996	437.43
0.998	430.14
0.999	414.74
0.9995	391.10
0.9999	528.11
0.99999	577.91

n [48]. This research also uses these stopping conditions with a maximum of 50,000 iterations ($m = 50,000$) and no more than 300 consecutive iterations without improvement ($n = 300$).

4.1.5 Simulated Annealing Algorithm

Given the discussion of components and parameters, Algorithm 2 summarizes the proposed simulated annealing heuristic steps. It is adapted from [42] to fit the parameters and notation already defined for this research. After initialization of the simulation, the main loop in the process (line 3) allows subsequent iterations until one of the stopping conditions is met. The bulk of simulation time occurs in lines 4 and 5, where a neighbor is chosen and then 1,200 simulation replications are conducted to determine $f(W^k)$. Lines 8 and 9 are required because it is possible that the most recently accepted network, W^A, is not optimal if line 14 accepted it in a prior iteration. The algorithm concludes by returning the best network structure identified and its respective average informed time.

4.1.6 Simulation Principles and Experimental Design

It is now important to focus on the framework of the simulation to ensure that the results are accurate, are replicable, and provide the information needed to make useful conclusions. This key component to simulation is experimental design, which aims to find the most impactful factors in a system with the least amount of simulation. The design used in this thesis is based on the principles described in detail in [48].

There are two main elements of experimental design, *factors* and *responses*. Factors are adjustable aspects, or inputs, of the simulation, while responses are the measurable statistical output of the simulation. Examples of factors in this simulation are the weight matrix, W^k, of the network, which, in turn, informs the rate matrix, R, as well as all adjustable parameters, like the number of replications per inject. The main

---
Algorithm 2 Simulated Annealing Process

1: Start with feasible network, W^0, and with average informed time, $f(W^0)$. The current optimal network $W^* = W^0$ and the most recently accepted network $W^A = W^0$. Iteration $k = 1$. $m = 50,000$ and $n = 300$. Number of Consecutive Non-Improving Iterations, $cni = 0$.
2: Determine a starting temperature $c = 0.9995$ and scaling factor $S = 100$.
3: **while** $k < m$ and $cni < n$ **do**
4:     Choose a random feasible neighbor, W^k, of current solution, W^A.
5:     Set $\Delta = f(W^k) - f(W^A)$
6:     **if** $\Delta \leq 0$ (improvement) **then**
7:         Set $W^A = W^k$.
8:         **if** $f(W^A) < f(W^*)$ **then**
9:             $W^* = W^A$.
10:         **end if**
11:         $cni = 0$.
12:     **else**
13:         Choose a u[0,1] random number, r.
14:         **if** $r < e^{-\Delta/Sc^k}$ **then**
15:             $W^A = W^k$.
16:         **end if**
17:         $cni = cni + 1$.
18:     **end if**
19:     $k = k + 1$.
20: **end while**
21: **return** W^* and $f(W^*)$
---

response for this simulation is the average informed time of the network. The built-in challenge of simulation is that the responses are only as accurate, and therefore, useful, as the experimental structure behind them.

When comparing the performance of various network structures in our simulation, it is critical to ensure that any changes in performance come from the difference in structure, and not simply as a result of different random numbers being used. This is where the importance of synchronizing the random number generator becomes a factor. In this simulation, there are five instances where a random number is required, abbreviated here as RV1,...,RV5, as in random variate. The first (RV1) is used to determine the transmission time for a node, as discussed in the previous chapter. The second instance (RV2) is used in conjunction with the R matrix to determine which node will receive the transmission of a particular node. The final three random variates relate to the network structure changes. Two random numbers (RV3 and RV4) are needed for determining which cell in W^k loses weight and which gains, respectively. The final random number (RV5) in the program is used to determine whether a network with worse performance than the previous network should be kept.

To reiterate, the best way to compare the performance of multiple network structures is to make the "randomness" in the simulation reproducible for all comparing structures. Law states, "*Ideally*, a specific

random number used for a specific purpose in one configuration is used for *exactly the same* purpose in all other configurations," [48] (original emphasis). Failure to accomplish this can cause insufficient and inaccurate statistical comparisons. For instance, if the first replication of inject 1 has the initial receiving node as **G** for one network structure but the first replication of the same inject for another structure has **I** as the first receiving node, then the change in network performance could be caused by this node difference, and not necessarily because of the network structure change.

The way to solve this problem is to establish a seed, or starting point, of a sequence of random numbers. This sequence, called a stream, will be the same as long as the same seed is used. Multiple streams are used in this research to account for the different random variates and ensure diversity in the random numbers. Using the same seed across the same replication ensures that the RV1 used to determine the first transmission time in a specific replication (say, replication 23) is the same RV1 used in replication 23 of all subsequent injects and iterations. Table 18 shows how each random variate is used, how many of them are needed, and if/when they get reset. Note that RV1 and RV2 each use different streams even though they require the same number and are reset at the same time. This is because they have two completely different purposes and pulling from two different streams of random numbers eliminates potential crossover of the two. This simulation uses unique streams of random numbers for RV1, RV2, and RV5. RV3 and RV4 (called RV3/4) can use the same stream because they are always called together and used for the same purpose, determining which cells gain and lose weight for W^k.

Table 18: Random Variate Characteristics

Random Variate	Usage	Number of Streams	When to Reset
RV1	Determining transmission time	400, one for each replication	Reset all 400 streams prior to next inject
RV2	Determining transmission recipient	400, one for each replication	Reset all 400 streams prior to next inject
RV3/4	Choosing cells to gain and lose weight in W^k	One stream	Only reset if testing multiple simulated annealing parameters in same simulation
RV5	Determining whether to accept a non-improving network structure	One stream	Only reset if testing multiple simulated annealing parameters in same simulation

The next challenge is in determining when to reset the seeds of these streams. There are really two sets of random variates in the simulation. RV1 and RV2 are used to run the simulation replications of the individual networks in line 5 of Algorithm 2, while RV3/4 and RV5 are used in the main simulated annealing

steps. Table 19 gives a visual representation of the structure of the simulation and when the various random number seeds are reset. The simulation begins with established base parameters for the simulated annealing. Then, it cycles through all required iterations of the simulated annealing. For each iteration, there are three injects, and for each inject, there are 400 replications. Before the start of every inject, all 400 seeds are reset for both RV1 and RV2. If different parameters are to be tested and compared without restarting the simulation, the seeds for RV3/4 and RV5 will need to be reset before restarting the simulated annealing process. An example for this use is in Section 4.1, which describes how this thesis determines the best cooling parameters to use.

Table 19: Visual Structure of the Simulated Annealing Heuristic

Base Parameters			
Iteration 0			
	Inject 1	Inject 2 (reset RV1 and RV2 seeds)	Inject 3 (reset RV1 and RV2 seeds)
	Replication 1	Replication 1	Replication 1
	Replication 2	Replication 2	Replication 2
	⋮	⋮	⋮
	Replication 400	Replication 400	Replication 400
Iteration 1			
	Inject 1 (reset RV1 and RV2 seeds)	Inject 2 (reset RV1 and RV2 seeds)	Inject 3 (reset RV1 and RV2 seeds)
	Rep 1 ... 400	Rep 1 ... 400	Rep 1 ... 400
⋮			
Final Iteration			
New Parameter (reset RV3/4 and RV5 seeds)			
Iteration 0 ... Final Iteration			

4.2 Conclusion

This chapter provides a detailed description of the proposed simulated annealing heuristic to identify an improving network structure for the FAD response network. Simulated annealing is a useful search heuristic for this application because it begins with a single network and allows for a broad range of potential results. The proposed version of simulated annealing is novel in that it uses discrete event simulation to calculate the objective function in each iteration. With the necessary parameters and experimental design factors in place, this model is ready for implementation to find a network structure that minimizes the network average informed time.

5 COMPUTATIONAL RESULTS

This chapter describes the results of applying the simulated annealing heuristic to the FAD exercise data to identify improved network structures. First, convergence characteristics and differences between neighborhood variations are discussed. This is followed by a discussion of the network with the best average informed time and the impactful changes in network structure. The chapter then concludes with a focus on the most significant relationships in the network and some general observations related to this FAD response exercise.

5.1 Simulated Annealing Process

The simulated annealing heuristic is implemented with four neighborhood variations. In implementing this simulation, the model runs the entire simulated annealing process five times with different seeds. In doing this, it ends up searching through five different paths of network changes, providing the best possible network structure over all of them in a reasonable amount of time.

Although the neighborhoods yield different rates of convergence to the final solutions, the quality of the solutions is quite similar. The differences in convergence for the first of the five simulated annealing seeds are illustrated in Figure 7, where the x-axis is total iterations and the y-axis is total improving iterations. Swap 3 converges to its best solution much faster than Swap 1, because Swap 1 makes smaller changes to the weights in each iteration. Swap 3, Swap 3-1, and Swap 3-2-1 are the same until around iteration 275, when 25 consecutive non-improving iterations have been completed. At this point, Swap 3-2-1 breaks off from Swap 3 and Swap 3-1. Then, around iteration 700, Swap 3-1 changes *swap* from three to one, while Swap 3-2-1 moves from two to one. From there, Swap 3-2-1 and Swap 3-1 progress together for the most part while Swap 1 and Swap 3 continue on their own. In the end Swap 1 has more than two and a half times more improving iterations than Swap 3 while Swap 3-1 and Swap 3-2-1 each have almost twice as many as Swap 3.

So how does this convergence time affect the quality of the network structure? Interestingly, the differences are quite minimal. A similar graph in Figure 8 shows how the neighborhood variations converge based on their optimal average informed time, $f(W^*)$. To make it possible to see differences, $ln(f(W^*))$ is shown. There is actually little variation between the neighborhoods, with the exception being Swap 1 from iterations 230-1,500. In the end, however, the four neighborhood variations produce network structures with very similar average informed times, as shown in Table 20. Another important note is that the steep drop in
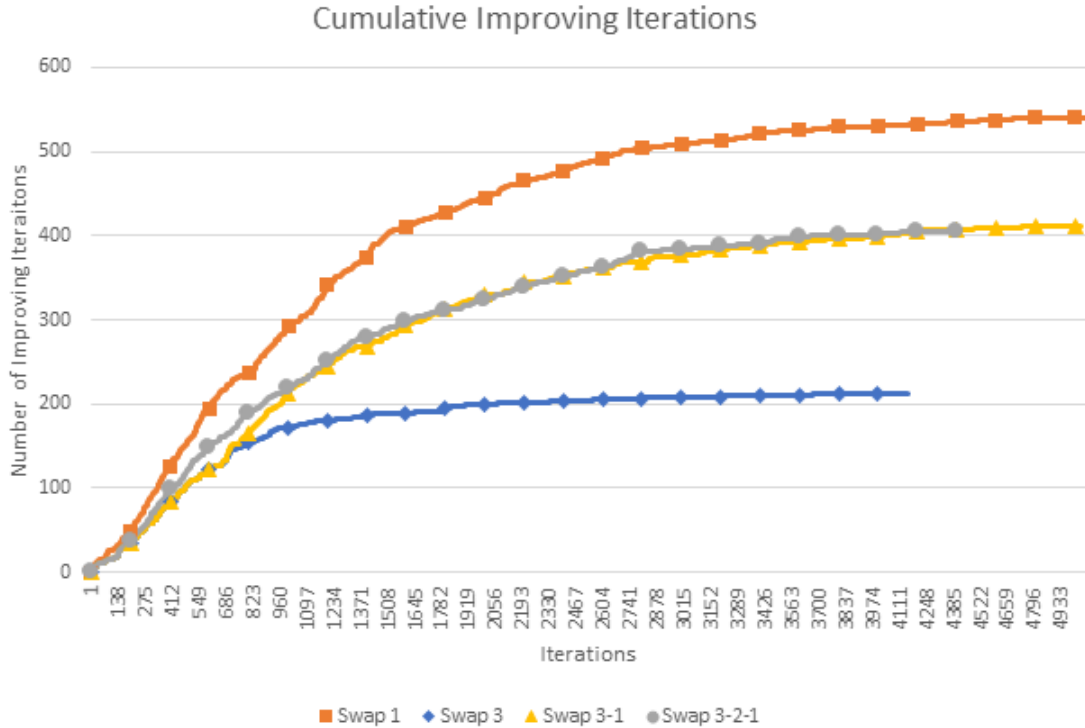
Figure 7: Total Number of Improving Iterations for Each Neighborhood Variation

the graph in Figure 8 around iteration 200 indicates that there are certain changes in the network structure that have particularly large impacts on the network's performance. In summary, Swap 3-2-1 reaches a better solution than Swap 1 in 68% as many iterations.

Table 20: Optimal Average Informed Time

	Swap 1	Swap 3	Swap 3-1	Swap 3-2-1
$f(W^*)$(min)	421.13	424.97	419.93	417.48
Optimal Iteration	5829	3444	5417	3975

The running time for the simulated annealing heuristic is similar for each of the network variations. The running time of the model is based on the number of iterations conducted; it completes about five iterations per second of run time, meaning the simulated annealing process described here takes around 92 minutes to complete depending on the neighborhood variation.

5.2   Most Improved Network

This section focuses on the best information dissemination network structures identified using the proposed simulated annealing heuristic. Specifically, it compares the best networks from each neighborhood
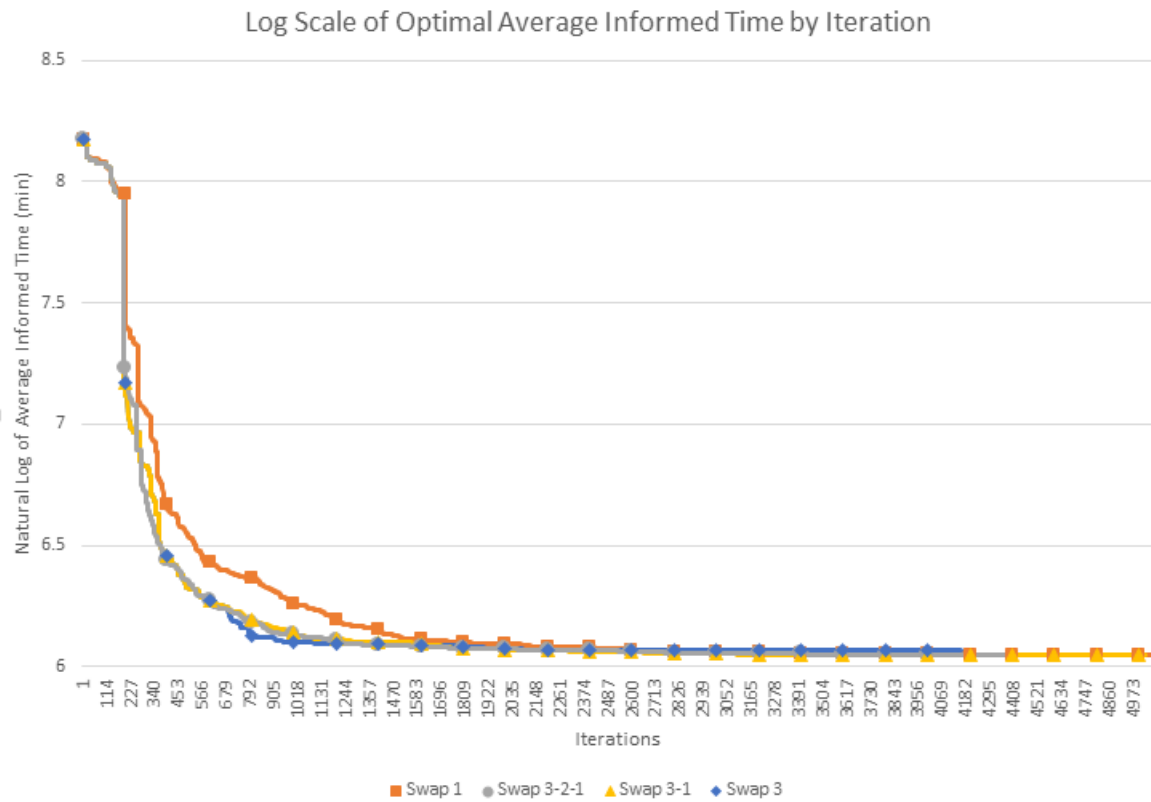
48

Figure 8: Log Scale of the Best Average Informed Time for Each Neighborhood Variation over the Number of Simulated Annealing Iterations

variation with each other and the base network and analyzes the importance of individual arcs on network performance.

5.2.1 Comparison to Base Network

After conducting five unique simulated annealing processes for all four neighborhood variations, the best network structure came from Swap 3-2-1 with an average informed time of 417.48 minutes. The resultant best network structure for that neighborhood, W^{*D}, is shown in Figure 9 in the same format as the base network, W^0, from Figure 4.

In comparing the two network structures from their respective figures, a few observations are apparent, many of which relate to the observations made regarding W^0 in 3.2.2.

- There are fewer total arcs in W^{*D} than W^0. In other words, it was often better for nodes to limit the number of nodes they communicate with and focus on those interactions that benefit the network most.

- KDA loses almost all of its incoming links. This is discussed further in Section 5.2.4.

- There is a large increase in communication between the county/local nodes, following Chapter 3, which predicted that would be an impactful change.

- The County/Local Administration and Kansas State EOC nodes established bi-directional communication after having no communication in W^0.

- The Federal, Local Vets, and Individual Producer nodes all dramatically increase their inbound communication. These were four of the last nodes to become informed in the W^0.

- The Unknown node loses most of its incoming communication. This is expected and the reason that the capacity is set at one, to shift communication to where it is better suited for the network.

5.2.2 Comparing Neighborhood Variations

With these general observations noted regarding W^{*D}, it is important to also look at the best networks found in the other neighborhood variations to determine if there are any consistent trends that all follow. A reasonable hypothesis for the closeness in $f(W^*)$ for the neighborhood variations, as shown in Table 20, is that they all have a similar optimal network. While there are certain arc changes that are similar across all four variations, denoted by the black arcs, Figure 10 shows that there are also quite a few differences in the optimal network structures of the neighborhood variations. For instance, Swap 3-1 is the only neighborhood variation that adds significant weight to the arc from Movement Control Branch to County/Local Emergency Management. Also, all variations show an increased weight to News Media, but the way nodes from which that weight is added varies.

The lines in this figure represent changes in each of the best network structures (W^{*A}, W^{*B}, W^{*C}, and W^{*D}) from the base W^0 by at least eight or more. A solid arc from node i to node j represents $w^*_{ij} \geq w^0_{ij} + 8$ while a dashed arc means $w^*_{ij} \leq w^0_{ij} - 8$. The thicker the arcs, the more weight is changed following the legend in the figure. The presence of black arcs, which indicate similar weight change for all neighborhood variations, are not surprising. Given that $f(W^*)$ for each of the neighborhood variations are within eight minutes of each other, one might expect far more similarities between them. The differences in arcs, coupled
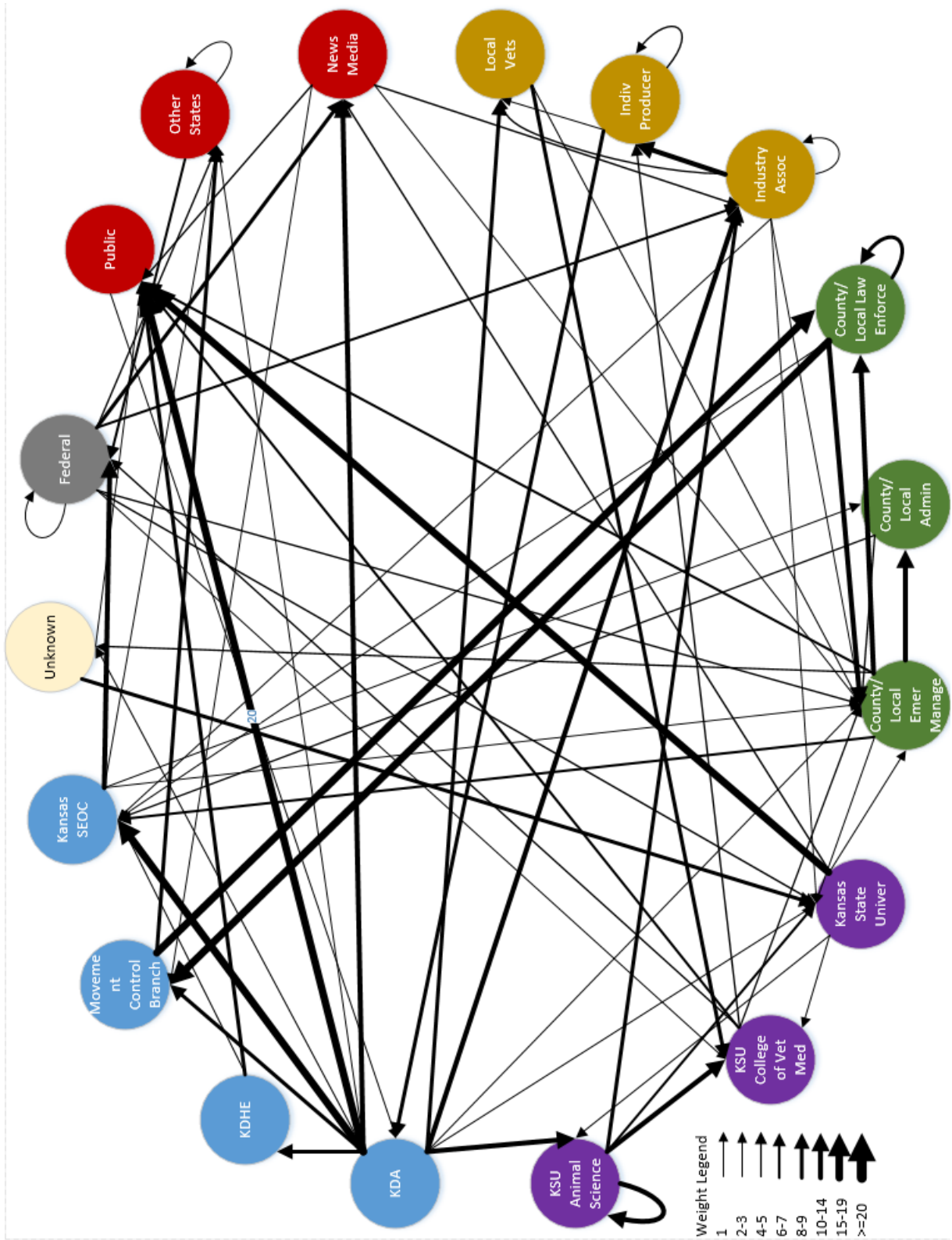
Figure 9: Flow Chart Representation of the Optimal Network after Simulated Annealing

with the closely related $f(W^*)$ of the variations seems to indicate that the black arcs are far more impactful to the overall performance of the network than others. This graph is a general means of comparison to other similar graphs, however, and does not show numerically how these arc changes impact the average informed time of the network.

5.2.3 Important Arc Changes

Up to this point there is no way to definitively tell which arc changes are most important to network structure improvement. To determine how various arc changes affect network performance, the changes in average informed time, Δ, are tracked for every iteration of the simulated annealing, regardless of whether the new network is accepted or not. This requires some additional variables and notation. Let Δ_{ij}^+ be the cumulative Δ for all iterations where w_{ij} gains link weight and Δ_{ij}^- be the cumulative Δ for all iterations where w_{ij} loses link weight. Then for each iteration in which w_{ij} increases, Δ_{ij}^+ is updated to $\Delta_{ij}^+ + \Delta$. Likewise, for each iteration in which w_{ij} decreases, Δ_{ij}^- is updated to $\Delta_{ij}^- + \Delta$. Similarly, let $swap_{ij}^+$ be the cumulative $swap$ for all iterations where w_{ij} gains link weight and $swap_{ij}^-$ be the cumulative $swap$ for all iterations where w_{ij} loses link weight. Then for each iteration in which w_{ij} increases, $swap_{ij}^+$ is updated to $swap_{ij}^+ + swap$. Likewise, for each iteration in which w_{ij} decreases, $swap_{ij}^-$ is updated to $swap_{ij}^- + swap$. After all iterations are complete, the average change in objective function when w_{ij} loses weight is $E(\Delta_{ij}^-) = \dfrac{\Delta_{ij}^-}{swap_{ij}^-}$ while that for w_{ik} gaining weight becomes $E(\Delta_{ik}^+) = \dfrac{\Delta_{ik}^-}{swap_{ik}^+}$. The result is a matrix with two cells for every cell in W^*, $E(\Delta_{ij}^-)$ and $E(\Delta_{ij}^+)$.

This matrix shows how specific network structure changes improve or degrade the network. Of all 648 cells in the matrix, only 18 have $E(\Delta_{ij}^-) < 0$ or $E(\Delta_{ij}^+) < 0$, meaning these changes are more likely to reduce the average informed time of the network. Sometimes improvements result from adding weight to an arc, while other instances illustrate that decreasing weight can improve average network informed time. Table 21 shows the 14 improving arc changes in green. It also includes the average objective function increase for when the opposite action occurs in orange. The difference between the two reveals the significance of an arc, because it shows how important one action is to improving the network as well as how detrimental the opposite action is on the network. The rows of Table 21 are sorted in non-increasing order of this time difference. Consider the first network change. Increasing the weight of the arc from Industry Associations to Individual Producers improves the average informed time by an average of 6.12 minutes for every additional unit of weight. Decreasing weight degrades network performance by an average of 10.33 minutes per arc
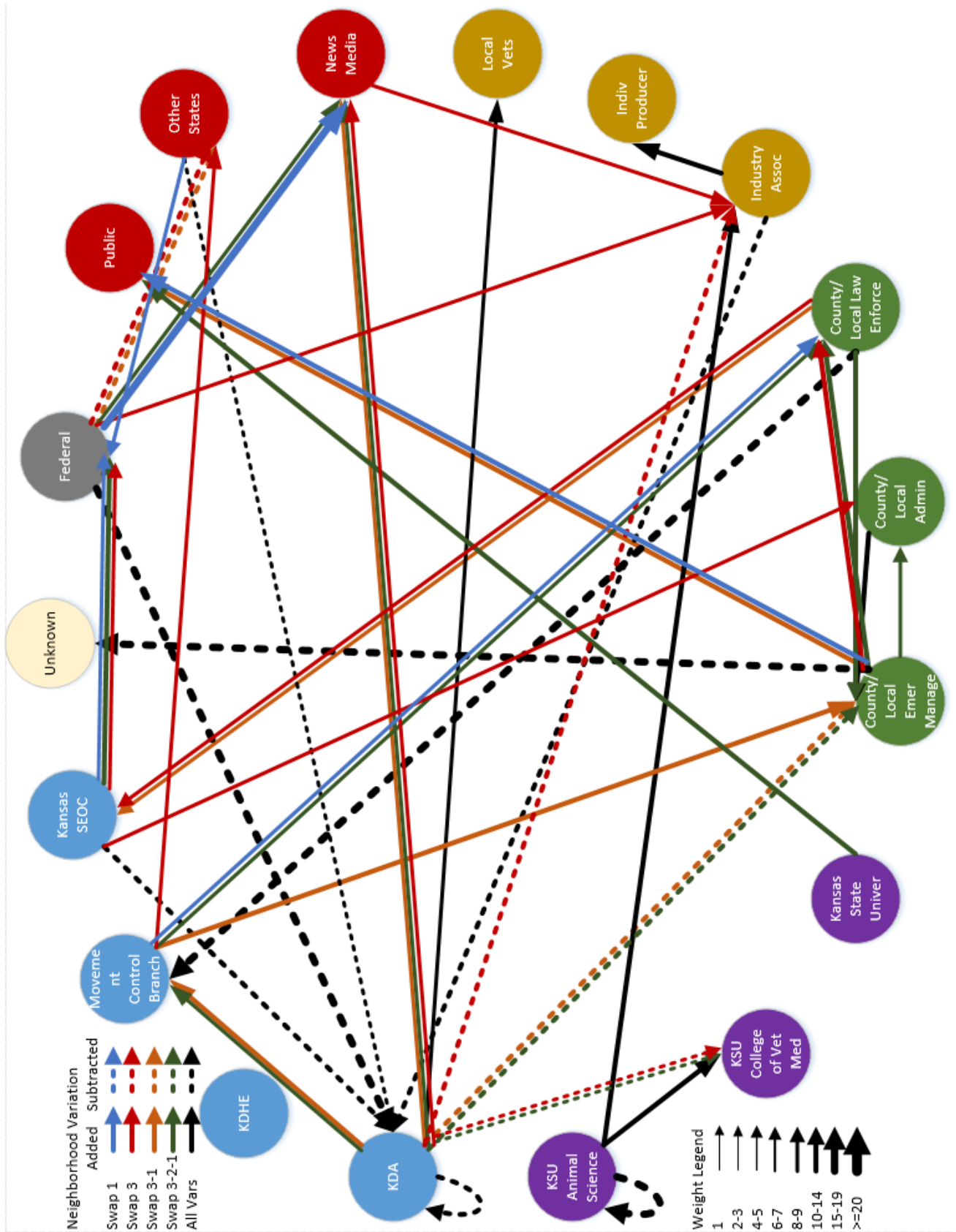
Figure 10: Flow Chart Showing the Most Significant Changes from the Base Network Structure for All Neighborhood Variations

weight. This difference is 6.77 minutes higher than the next best, showing the importance of the relationship between these two nodes.

The right side of this table shows how the network for each of the neighborhood variations takes advantage of these important arcs. The four right-most columns summarize the change in weight for the arc (i,j) between the nodes specified in each row, comparing the final solution weight w_{ij}^* and the initial weight w_{ij}^0. The only two improving network changes not used are KDHE to KDA and Industry Associations to County/Local Emergency Management. KDHE has no weight going to KDA in W^0, so using that improving delta is impractical. For the arc from Industry Associations to County/Local Emergency Management, all neighborhood variations increase weight instead of decrease. This is because of the two improving changes directly below in the table. The average total improvement of decreasing Industry Associations to itself and Industry Associations to KDA is greater than that of decreasing Industry Associations to County/Local Emergency Management. Taking full advantage of the former two requires adding some weight to the latter.

In the end, this table shows how the most impactful links are altered among the four neighborhood variations. It also shows why Swap 3-1 (W^{*C}) and Swap 3-2-1 (W^{*D}) outperformed Swap 1 (W^{*A})and Swap 3 (W^{*B}). W^{*C} and W^{*D} both utilize the top three average improving deltas better than W^{*A} and W^{*B} do. This is at least part of the reason that these network structures perform better.

5.2.4   KDA Change Analysis

Of all the observations about the best network structures produced by simulated annealing, the most obvious, most impactful, and potentially most concerning is the shift of nearly all arcs away from KDA($\mathbf{I}$). The total incoming weight to KDA ($\sum_{i=1}^{n} w_{i,\mathbf{I}}$) is only eight or nine in the solution produced by each neighborhood, compared to 111 in W^0. Given that KDA is the coordinating agency for the FAD response, eliminating nearly all incoming communication could appear contradictory to the goals of this research. It is important to consider, however, that this simulation equates all information passing as equal and there is no benefit for a node to receive information after it is already informed. This means that once KDA has reached its capacity, it no longer needs to receive anything. Its relatively low λ_i^{max} compared to its outgoing links makes this surprising change more understandable.

This drastic change in KDA does support a main observation from Chapter 3, that KDA may not have sufficient capacity to effectively manage all its communication demands. It serves as an information broker to the rest of the network but does not seem to have enough personnel to pull in communication from other nodes.

Table 21: Network Structure Changes That Create a Negative Average Change in Average Informed Time of the Network

Sending Node (i)	Receiving Node (j)	Change in Average Informed Time (min)			$w_{ij}^* - w_{ij}^0$ for each Neighborhood Variation			
		$E(\Delta_{ij}^+)$	$E(\Delta_{ij}^-)$	Time Diff.	Swap 1	Swap 3	Swap 3-1	Swap 3-2-1
Industry Associations	Individual Producers	-6.12	10.33	16.45	8	9	12	10
KSU College of Vet Medicine	KDA	5.36	-4.32	9.68	-3	-5	-4	-5
County/Local Emergency Management	County/Local Administration	-0.4	8.67	9.07	4	2	6	9
KSU College of Vet Medicine	Unknown	7.9	-0.7	8.6	-1	-1	-1	-1
Industry Associations	County/Local Emergency Management	7.94	-0.16	8.1	2	1	1	1
Industry Associations	Industry Associations	5.45	-0.51	5.96	-3	-3	-3	-2
Industry Associations	KDA	3.05	-2.47	5.52	-14	-14	-14	-14
Federal	Federal	-0.68	2.68	3.36	3	3	4	3
Other States	KDA	1.48	-1.7	3.18	-9	-9	-9	-9
Other States	Federal	-0.68	2.315	2.995	8	6	6	6
Kansas State EOC	KDA	2.36	-0.41	2.77	-11	-11	-11	-11
KDHE	KDA	1.12	-0.49	1.61	0	0	0	0
MCB	KDA	1.11	-0.13	1.24	-5	-7	-5	-5
County/Local Law Enforcement	KDA	1.01	-0.15	1.16	-1	-1	-1	-1

Table 22: Effects of KDA Capacity on W^{*D} for One Simulated Annealing Instance

KDA Max Capacity	Total Incoming Weight	$f(W^*)$(min)
100	112	464.3
80	94	454.76
60	69	446.06
40	47	431.79
30	33	423.87
20	16	423
13	9	417.48

To further examine this, the maximum capacity for KDA is increased to examine the relationship between capacity, total incoming weight, and average network informed time. The Swap 3-2-1 neighborhood is used and one simulated annealing run is executed for KDA maximum capacity values ranging from 13 (the value of the original network) to 100. Table 22 summarizes the results of this testing and shows a positive correlation between KDA's maximum capacity, the total incoming weight, and the average informed time for W^*. KDA's total incoming weight in W^0 is 111 with an average informed time of 3,570.31 minutes. With a capacity of 100, the total incoming weight is nearly identical (112), yet the average informed time is close to that of $f(W^*)$ with the original capacities. Another way to interpret this result is that with the same amount of total incoming weight and the right configuration in the rest of the network, KDA needs a capacity of 100 to approach a similar average informed time as was achieved in the optimal solutions.

5.3   Conclusion

The results described in this chapter demonstrate that simulated annealing is an effective way to improve the network structure of this FAD response exercise. By using different variations of neighborhoods, a collection of strong performing networks are created, with the best network structure coming from Swap 3-2-1. This network structure also decreases the network average informed time from 3,500 minutes in W^0 to just 417 minutes. This best network also yields some useful observations about the network as a whole when compared with the base network structure.

The collection of best networks for the neighborhood variations allows for further analysis of the arcs themselves to determine which have the most significant impact on network performance. This chapter shows that each of these neighborhoods use the most important arc changes to some degree, but the progressive neighborhoods (Swap 3-1 and Swap 3-2-1) use them better than Swap 1 and Swap 3. The results in this

chapter provide a strong foundation from which to draw conclusions and give recommendations for future

FAD responses.

6 CONCLUSIONS AND FUTURE WORK

The research described in this thesis leads to specific recommendations and avenues for future research. This chapter discusses these areas.

6.1 Recommendations

Recommendations for improvement are divided into three categories: FAD preparedness at KDA and similar agencies, for other researchers in the field of crisis communication, and for implementation of this model in other forms. Each will be discussed in turn.

6.1.1 Kansas Department of Agriculture

This research reveals a few areas that can be addressed to improve the communication surrounding future FAD response exercises or events. Because KDA is the synchronizing organization, these recommendations are primarily directed toward emergency response personnel there, though some also apply to other stakeholders in such events.

First of all, the best networks found by simulated annealing show the KDA node losing 92% of its incoming communication links. The discussion in 5.2.4 shows that increasing the capacity of the KDA node helps maintain the incoming information flow to the node while simultaneously keeping a near-optimal network informed time. This confirms some of the earlier observations in 3.2.2 that KDA should increase its capacity to effectively manage information demand across the network and the processing of incoming information.

Another revealing factor in this research is the relationship between the various county/local nodes and industry-related nodes. Increasing communication between these two groups of nodes significantly improves information flow in the network, as all of the best networks have significant increase in the weight among these relationships.

Another tangible recommendation for future exercises and event responses is to improve the data collection process as outlined in 3.4. Adding moderators and requiring participants to log communications will help develop a more accurate network for further analysis. Also, expanding the participation in the exercise to include as many other important groups (industry associations, checkpoint operators, ranchers, local veterinarians, etc.) as is feasible will improve the quality of the data immensely.

6.1.2 Other Researchers

For other researchers in the field of information dissemination or crisis communication, one important application within this research is the idea of clustering of players to develop a more usable communication network. Dealing with empirical data inevitably leaves gaps and inconsistencies in the model. This is likely a reason for the lack of applied research in this field. This idea connects with how stakeholders are represented in such a model. Clustering these groups together and aggregating their data allows for further analysis of the network without compromising the integrity of the model.

6.1.3 Model Implementation

For others interested in implementing this model with other data sets, the progressive shift Swap 3-2-1 is the recommendation for future implementation of the model. Section 4.1.1 defines the neighborhood variations used for the simulated annealing process. The average informed times from those variations were quite close, but the progressive shift Swap 3-2-1 performed best.

6.2 Future Work

There is a lot of potential for future work in the area of crisis communication networks, particularly with regards to quantitative methods. Building on the research in this thesis, the future work described in this section has the potential to yield additional improvements in crisis communication network design.

The first area of future research involves changes in the neighborhood definition and/or objective function value used in the simulated annealing heuristic. For example, the neighborhood could be based on increased or altered capacities, not just arc weights. While average informed time is an appropriate objective function for this context, other FAD response goals should also be explored in the heuristic framework. These include time to inform priority nodes or priority-weighted informed time depending on the type of information being shared (corresponding to different injects).

The second area of future research involves extending the network model itself to account for additional characteristics of a FAD response. For example, the modes of communication transmission (i.e. phone conversation, text, press release) may play a role in the performance of the network. In addition, through discussions with KDA, it is clear that relationships and trust between different pairs of nodes vary. This is probably the most interesting and potentially significant next step with this research. There has been work related to trust in the literature [18, 39, 51, 52, 53] but nothing that explores it quantitatively as it relates

to networks. Specifically, an interesting next step is identifying whether the amount of trust one has in the information he receives based on the sender impacts the best possible network. Another avenue of research involves some of the assumptions made in this simulation model. The model could be modified such that information does not necessarily need to reach every node or capacity in the network. Additionally, this model assumes that one transmission comes from one node to another specific node but certain modes of communication, especially email, could send the same information to many nodes at the same time. This is not depicted in the current model but leaves room for future work.

7 APPENDIX I - CENTRALITY

This appendix serves to describe in detail network centrality, how this thesis addressed its potential use, and why it does not accurately predict average informed time of the network.

As was discussed in Chapter 1, centrality is one of the most commonly used metrics that help classify and characterize networks. Freeman develops the three primary types of centrality in his seminal work on the subject [16]: degree centrality, closeness centrality, and betweenness centrality. Degree centrality relates to the number of links incident to a particular node and is considered a local centrality metric, meaning it only takes into account the most immediate neighbors of a node and not the network at large. Closeness and betweenness centrality are considered global metrics in that they consider the entirety of the network. Both are calculated using the shortest paths between nodes. The closeness centrality (C_C) of a node is equal to the sum of its shortest paths to all other nodes in the network. The betweenness centrality (C_B) of a node is the number of times that node is included in the shortest paths of all other nodes to all nodes. As an example, if we assume all arcs to have a weight of 1 in Figure 11, Table 23 shows the shortest paths for all nodes.
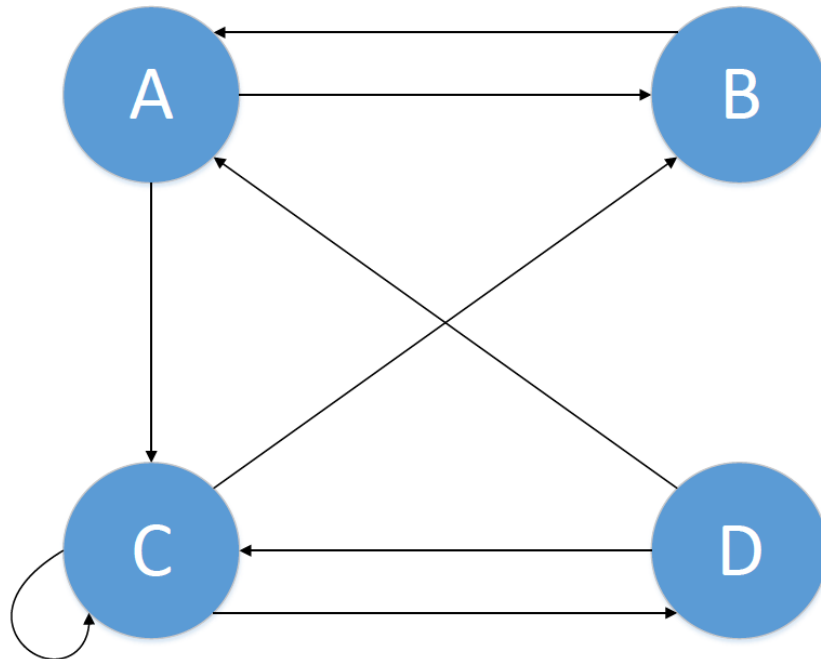

Figure 11: Diagram of Example Network

Node A in this network has a C_C of 4 because the shortest paths from A to B and A to C are 1 arc long, while the shortest path from A to D is 2 arcs long. The C_B of A is 5.5 because A is on the singular shortest path of 4 node combinations (those originating at A do not count towards C_B). It is also on both shortest paths from C to A, thus gaining a value of 1, but only on one of the two shortest paths from D to B, gaining an additional .5. Note that the sum of all C_C values equals the sum of all C_B, just distributed differently. This reveals the underlying difference of the two metrics.

Table 23: Shortest Paths of Example Network

Shortest Paths	A	B	C	D	C_C	C_B
A	0	A,B	A,C	A,C,D	4	5.5
B	B,A	0	B,A,C	B,A,C,D	6	3.5
C	C,B,A & C,D,A	C,B	0	C,D	4	5.5
D	D,A	D,A,B & D,C,B	D,C	0	4	3.5

[54, 55, 34] have all pursued centrality in weighted networks to determine how the various measures relate to each other and the networks they represent. Borgatti looks at many types of networks, two of which closely resemble our information dissemination network. His Gossip and Infection networks have very similar properties in that they both represent network flow via serial duplication. This means transmission is conducted one at a time(serial) as opposed to many nodes at once (parallel) and that the gossip/infection is not lost by the sending node but duplicated before transmission. He states that low closeness centrality scores in information flow networks tend to have better access to new information, indicating that closeness centrality could be our best option for statistical prediction of network performance. He further found that closeness centrality is the best known centrality measure for geodesic serial duplication (taking the shortest possible route between two nodes). However, his work was inconclusive as to an appropriate measure for path and trail serial duplication which represent the infection and gossip networks, respectively. The lack of a confirmed best centrality statistic leads us to pursue closeness centrality over the other options based on the general theoretical guidelines of the statistic his research provides.

[56] developed methods to assign centrality metrics to weighted networks. In a weighted network, a higher weight means a stronger or "closer" link from one node to another. However, to find C_C and C_B requires the shortest paths between nodes, something that will be inaccurate in a weighted network because it will always choose the weaker of the links. In order to find the shortest paths, we must take the inverse of all the weights before applying a shortest path algorithm. Figure 12 shows how the example network would look

with inverse weights. From here we can calculate our centrality measurements as in Table 24. In a weighted network, unlike the previous example, the sum of the C_C does not necessarily equal the sum of the C_B.

With this information, we can now determine a centrality metric that encapsulates the network as a whole. Because we are using closeness centrality in our model, we can disregard betweenness. Freeman offers two such metrics, average and dominance of the network[16]. Average is simply the average centrality measure of all nodes. In this case, $\dfrac{1\frac{1}{12} + 1\frac{5}{6} + 1\frac{5}{6} + 1\frac{5}{12}}{4} = 1\dfrac{13}{24}$. Dominance($C_C^{dom}$), however, measures how strongly one node dominates the activity in the network by comparing the highest centrality value (C_C^{max}) in the network to that of all other nodes. The equation for this value and its result in this example is
$$C_C^{dom} = \frac{\sum_{i=1}^{n}(C_C^{max} - C_C^i)}{n-1} = \frac{\frac{3}{4} + 0 + 0 + \frac{5}{12}}{3} = \frac{7}{18}$$
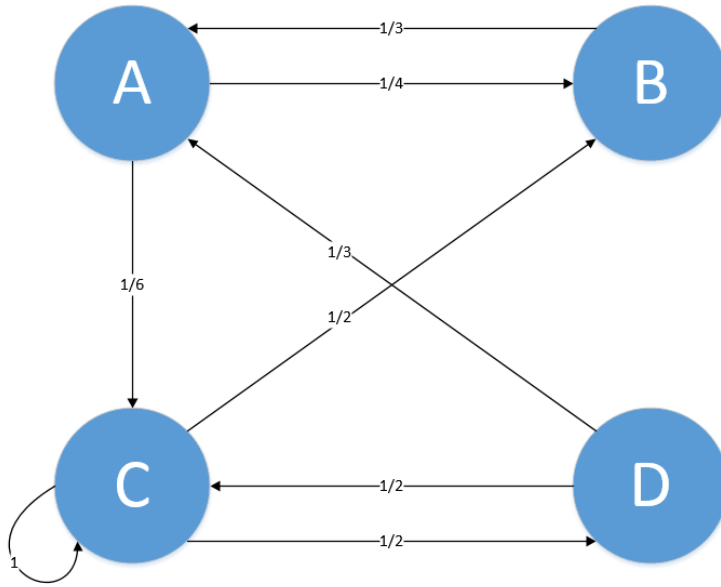

Figure 12: Inverse Weighted Example Network

Table 24: Shortest Paths of Example Inverse Weighted Network

Shortest Paths	A	B	C	D	C_C	C_B
A	0	A,B ($\frac{1}{4}$)	A,C ($\frac{1}{6}$)	A,C,D ($\frac{2}{3}$)	($1\frac{1}{12}$)	6.5
B	B,A ($\frac{1}{3}$)	0	B,A,C ($\frac{1}{2}$)	B,A,C,D (1)	($1\frac{5}{6}$)	3.5
C	C,B,A & C,D,A ($\frac{5}{6}$)	C,B ($\frac{1}{2}$)	0	C,D ($\frac{1}{2}$)	($1\frac{5}{6}$)	5
D	D,A ($\frac{1}{3}$)	D,A,B ($\frac{7}{12}$)	D,C & D,A,C ($\frac{1}{2}$)	0	($1\frac{5}{12}$)	3.5

One strategy considered in this research, but not ultimately used, is that of utilizing network centrality to determine how to alter the network structure. Closeness centrality is a network statistic that considers the entire network and its weights and calculates how close (via the weighted arcs in the network) one node is to every other node [16]. A more thorough explanation of the different kinds of centrality and an example for calculating closeness centrality can be found in Appendix I.

The goal in finding these closeness centrality measures is to be able to adjust the network structures based on their centrality in order to improve the average informed time. However, before implementation, a correlation between network centrality and average informed time must be verified.

To do this, 4 test structures are considered, each with varying levels of centrality, to determine their average informed time. The details of the test structures are:

- Test 1 - Increase the lowest five KDA outgoing weights by 3 and decrease the greatest five by 3.

- Test 2 - Add a weight of 2 to all cells with a weight of 0 or 1 whose respective cell in the A matrix is 1.

- Test 3 - Complete graph with links to and from all nodes, keeping the same total number of links from the base structure and dividing them equally.

- Test 4 - Star graph with KDA as the center, keeping the same total number of links from the base structure and dividing them equally.

Figure 13 shows how the average informed time of the various test structures compares to their respective centrality measurements. Average centrality is simply the average of all nodes, maximum and minimum centralities represent the most and least central node measurements, respectively, and the dominance is the cumulative difference in centrality value of all nodes from the most central node. While it appears from Test 3 and 4 that centrality may have a positive correlation with average informed time, the comparison of the base structure shows that to not necessarily be the case. In short, there is not enough evidence to show that centrality is an accurate predictor of average informed time, and, therefore, it is not used in this model.
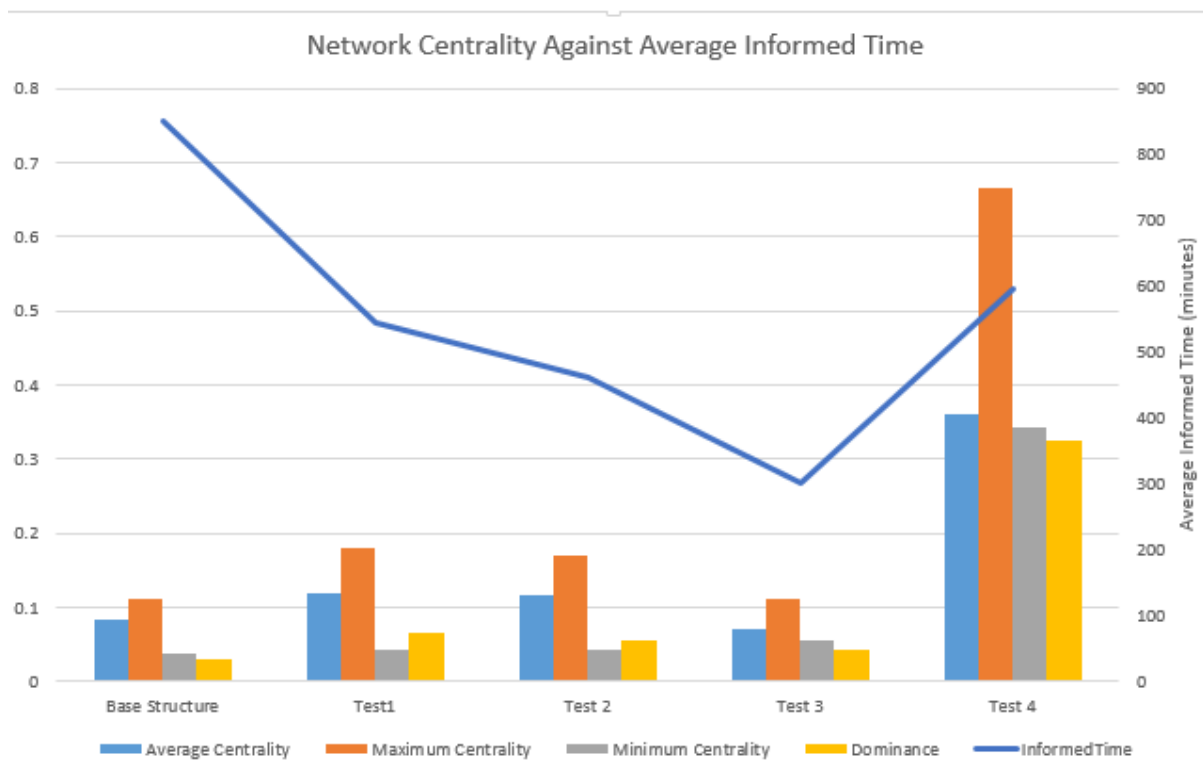
Figure 13: Network Centrality Against Average Informed Time

8 APPENDIX II - DATA TRANSLATION EXAMPLE

This appendix serves to give a more thorough explanation of the how the extracted data translates into the network structure. After extracting the raw data and organizing it in the Extracted Transmission List, the next step is to develop a model from this list. Specifically, this involves using the information from each of the 431 transmissions to complete the weight matrix, W, and the capacities for each node, all initialized to 0. In general, for every transmission from node i to node j, $w_{ij} = w_{ij} + 1$. Also, for every player not yet seen in the Extracted Transmission List for node i, $\lambda_i^{max} = \lambda_i^{max} + 1$. The example set of three transmissions found in Table 25 is used here to demonstrate this process.

Table 25: Extracted Transmission List Example

Time	From Cluster	From Player	To Cluster	To Player
9:23	KDA (**O**)	Kansas State Veterinarian	Industry Associations (**F**)	Midwest Dairy Farmers
9:25	KDA (**O**)	Kansas State Veterinarian	Industry Associations (**F**)	Kansas Zoological Parks
9:45	Industry Associations (**F**)	Midwest Dairy Farmers	KDA (**O**)	KDA Animal Health

Table 26: Extracted Transmission List Example Results

After Time	$w_{\mathbf{OF}}$	$w_{\mathbf{FO}}$	$\lambda_{\mathbf{O}}^{max}$	$\lambda_{\mathbf{F}}^{max}$
9:23	1	0	1	1
9:25	2	0	1	2
9:45	2	1	2	2

Assuming that these are the first three transmissions in the Extracted Transmission List (e.g. $w_{ij} = 0$ for all $i, j \in \{N\}$ and $\lambda_i^{max} = 0$ for all $i \in \{N\}$), Table 26 shows how the weight matrix and capacities are filled after each transmission. Notice that after the 9:25 transmission, the $\lambda_{\mathbf{O}}^{max}$ remains 1 because the Kansas State Veterinarian has already been represented for node **O** during the 9:23 transmission. Similarly, $\lambda_{\mathbf{F}}^{max}$ stays the same after the 9:45 transmission even though the Midwest Dairy Farmers player is now the sending player instead of receiving as in the 9:23 transmission.

9 BIBLIOGRAPHY

References

[1] B. Lee, J. Park, P. Gordon, J. Moore, and H. W. Richardson. Estimating the state-by-state economic impacts of a foot-and-mouth disease attack. *International Regional Science Review*, 35(1):26–47, 2012.

[2] Animal and Plant Health Inspection Service. Foreign animal disease. Technical report, United States Department of Agriculture, Jul 2013. Safeguarding American Agriculture.

[3] Animal and Plant Health Inspection Service. Foot-and-mouth disease. Technical report, United States Department of Agriculture, Jul 2013. Safeguarding American Agriculture.

[4] T. E. Carpenter, J. M. O'Brien, A. D. Hagerman, and B. A. McCarl. Epidemic and economic impacts of delayed detection of foot-and-mouth disease: A case study of a simulated outbreak in California. *Journal of Veterinary Diagnostic Investigation*, 23(1):26–33, Jan 2011.

[5] R. R. Kao, D. M. Green, J. Johnson, and I. Z. Kiss. Disease dynamics over very different time-scales: Foot-and-mouth disease and scrapie on the network of livestock movements in the UK. *Journal of the Royal Society Interface*, 4(16):907–916, Oct 2007.

[6] Kansas Department of Agriculture. Food and agriculture incident annex. Technical report, 2012. Kansas Response Plan.

[7] H. Yoon, S. S. Yoon, H. Kim, Y. J. Kim, B. Kim, and S. H. Wee. Estimation of the infection window for the 2010/2011 Korean foot-and-mouth disease outbreak. *Osong Public Health and Research Perspectives*, 4(3):127, 2013.

[8] A. H. Delgado, B. Norby, H. M. Scott, W. Dean, W. A. McIntosh, and E. Bush. Distribution of cow−calf producers' beliefs about reporting cattle with clinical signs of foot-and-mouth disease to a veterinarian before or during a hypothetical outbreak. *Preventive Veterinary Medicine*, 117(3-4):505–517, 2014.

[9] A. H. Delgado, B. Norby, H. M. Scott, W. Dean, W. A. McIntosh, and E. Bush. Distribution of cow-calf producers' beliefs regarding gathering and holding their cattle and observing animal movement restrictions during an outbreak of foot-and-mouth disease. *Preventive Veterinary Medicine*, 117(3-4):518–532, 2014.

[10] M. S. Pishvaee, K. Kianfar, and B. Karimi. Reverse logistics network design using simulated annealing. *International Journal of Advanced Manufacturing Technology*, 47(1-4):269–281, Mar 2010.

[11] N. Chibeles-Martins, T. Pinto-Varela, A. P. Barbosa-Povoa, and A. Q. Novais. A multi-objective meta-heuristic approach for the design and planning of green supply chains - MBSA. *Expert Systems with Applications*, 47:71–84, Apr 2016.

[12] I. Samora, M. J. Franca, A. J. Schleiss, and H. M. Ramos. Simulated annealing in optimization of energy production in a water supply network. *Water Resources Management*, 30(4):1533–1547, Mar 2016.

[13] R. K. Ahuja, T. L. M., and J. B. Orlin. *Network Flows: Theory, Algorithms, and Application*. Prentice Hall, First edition, 1993.

[14] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America*, 97(21):11149–11152, Oct 2010 2000.

[15] A.L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, Oct 15 1999.

[16] L. C. Freeman. Set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.

[17] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):7821–7826, Jun 2002.

[18] J. C. Long, F. C. Cunningham, and J. Braithwaite. Bridges, brokers and boundary spanners in collaborative networks: A systematic review. *BMC Health Services Research*, 13:158, Apr 2013.

[19] J. Braithwaite. Between-group behaviour in health care: Gaps, edges, boundaries, disconnections, weak ties, spaces and holes. a systematic review. *BMC Health Services Research*, 10:330, Dec 2010.

[20] M. L. Popovich, J. M. Henderson, and J. Stinn. Information technology in the age of emergency public health response. *IEEE Engineering in Medicine and Biology Magazine*, 21(5):48–55, Sep 2002.

[21] R. K. Sharma, B. R. Gurjar, A. V. Singhal, S. R. Wate, S. P. Ghuge, and R. Agrawal. Automation of emergency response for petroleum oil storage terminals. *Safety Science*, 72:262–273, 2015.

[22] A. Seki, O. Saito, H. Nago, K. Suzuki, K. Tomishima, K. Saito, and H. Takemiya. Development of a software platform for providing environmental monitoring data for the Fukushima Daiichi nuclear accident. *Radiation Protection Dosimetry*, 164(1-2):97–102, Apr 2015.

[23] L. Yang, S. H. Yang, and L. Plotnick. How the Internet of Things technology enhances emergency response operations. *Technological Forecasting and Social Change*, 80(9):1854–1867, Nov 2013.

[24] Merriam-Webster. Merriam-Webster Dictionary. http://www.merriam-webster.com. Accessed: 6 April 2016.

[25] V. Slavkovikj, S. Verstockt, S. Van Hoecke, and R. Van de Walle. Review of wildfire detection using social media. *Fire Safety Journal*, 68:109–118, Aug 2014.

[26] E. Sjoberg, G. C. Barker, J. Landgren, I. Griberg, J. E. Skiby, A. Tubbin, A. von Stapelmohr, M. Harenstam, M. Jansson, and R. Knutsson. Social media and its dual use in biopreparedness: Communication and visualization tools in an animal bioterrorism incident. *Biosecurity and Bioterrorism-Biodefense Strategy Practice and Science*, 11:S264–S275, Sep 2013.

[27] J. B. Houston, J. Hawthorne, M. F. Perreault, E. H. Park, M. G. Hode, M. R. Halliwell, S. E. T. McGowen, R. Davis, S. Vaid, J. A. McElderry, and S. A. Griffith. Social media and disasters: A functional framework for social media use in disaster planning, response, and research. *Disasters*, 39(1):1–22, Jan 2015.

[28] L. Gao, C. Song, Z. Gao, A. L. Barabasi, J. P. Bagrow, and D. Wang. Quantifying information flow during emergencies. *Scientific Reports*, 4:3997, Feb 2014.

[29] J. L. Garnett and A. Kouzmin. Communicating throughout Katrina: Competing and complementary conceptual lenses on crisis communication. *Public Administration Review*, 67:171–188, Dec 2007.

[30] G. S. Leonard, C. Stewart, T. M. Wilson, J. N. Procter, B. J. Scott, H. J. Keys, G. E. Jolly, J. B. Wardman, S. J. Cronin, and S. K. McBride. Integrating multidisciplinary science, modelling and impact data into evolving, syn-event volcanic hazard mapping and communication: A case study from the 2012 Tongariro eruption crisis, New Zealand. *Journal of Volcanology and Geothermal Research*, 286:208–232, Oct 2014.

[31] E. Savoia, L. Lin, and K. Viswanath. Communications in public health emergency preparedness: A systematic review of the literature. *Biosecurity and Bioterrorism-Biodefense Strategy Practice and Science*, 11(3):170–184, Sep 2013.

[32] E. Savoia, S. B. Massin-Short, A. M. Rodday, L. A. Aaron, M. A. Higdon, and M. A. Stoto. Public health systems research in emergency preparedness: A review of the literature. *American Journal of Preventive Medicine*, 37(2):150–156, Aug 2009.

[33] K. Zhao, J. Yen, L. M. Ngamassi, C. Maitland, and A. H. Tapia. Simulating inter-organizational collaboration network: A multi-relational and event-based approach. *Simulation-Transactions of the Society for Modeling and Simulation International*, 88(5):617–633, May 2012.

[34] S. P. Borgatti. Centrality and network flow. *Social Networks*, 27(1):55–71, Jan 2005.

[35] D. Clayton and M. Hills. *Statistical Models in Epidemiology*. Oxford University Press, First edition, 1993.

[36] O. Diekmann and J. A. P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. John Wiley and Sons, First edition, 2000.

[37] M. Youssef and C. Scoglio. Mitigation of epidemics in contact networks through optimal contact adaptation. *Mathematical Biosciences and Engineering*, 10(4):1227–1251, 2013.

[38] H. Shakeri, F. D. Sahneh, C. Scoglio, P. Poggi-Corradini, and V. M. Preciado. Optimal information dissemination strategy to promote preventive behaviors in multilayer epidemic networks. *Mathematical Biosciences and Engineering*, 12(3):609–623, 2015.

[39] W. G. Kennedy. Modeling human behaviour in agent-based modeling. In A. J. Heppenstall, A. T. Crooks, L. M. See, and M. Batty, editors, *Agent-Based Models of Geographical Systems*, pages 167–179. Springer, 2012.

[40] T. A. Au. Analysis of command and control networks on Black Saturday. *Australian Journal of Emergency Management*, 26(3):20–29, 2011.

[41] J. Hamra, R. Wigand, L. Hossain, and C. Owen. Network effects on learning during emergency events. *Knowledge Management Research & Practice*, 12(4):387–397, Nov 2014.

[42] D. S. Johnson and L. A. McGeoch. The traveling salesman problem: A case study. In E. Aarts and J. K. Lenstra, editors, *Local Search in Combinatorial Optimization*, pages 216–309. Princeton University Press, 2003.

[43] P. Pezzini, O. Gomis-Bellmunt, and A. Sudria-Andreu. Optimization techniques to improve energy efficiency in power systems. *Renewable & Sustainable Energy Reviews*, 15(4):2028–2041, May 2011.

[44] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[45] J. C. Gower and G. J. S. Ross. Minimum spanning trees and single linkage cluster analysis. *Journal of the Royal Statistical Society*, 18(1):54–64, 1969.

[46] C. T. Zahn. Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Transactions on Computers*, C 20(1):68–86, 1971.

[47] C. Fraley and A. E. Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458):611–631, Jun 2002.

[48] A. M. Law. *Simulation Modeling and Analysis*. McGraw Hill, Fourth edition, 2007.

[49] Mathwave Technologies. Easyfit (version 5.6). http://www.mathwave.com/downloads.html.

[50] W. Mendenhall and T. Sincich. *Regression Analysis*. Prentice Hall, Seventh edition, 2012.

[51] M. Mejia, N. Pena, J. L. Munoz, and O. Esparza. A review of trust modeling in ad hoc networks. *Internet Research*, 19(1):88–104, 2009.

[52] D. S. Staples and J. Webster. Exploring the effects of trust, task interdependence and virtualness on knowledge sharing in teams. *Information Systems Journal*, 18(6):617–640, Nov 2008.

[53] B. J. Reynolds. When the facts are just not enough: Credibly communicating about risk is riskier when emotions run high and time is short. *Toxicology and Applied Pharmacology*, 254(2):206–214, Jul 2011.

[54] T. Opsahl, F. Agneessens, and J. Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245–251, Jul 2010.

[55] A. Barrat, M. Barthlemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11):3747–3752, Mar 2004.

[56] M. E. J. Newman. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 73(3):039906, Mar 2006.