

Looking at Community Issues Through the Lens of Mathematical Modeling

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Anna H. Sisk

May 2022

© by Anna H. Sisk, 2022
All Rights Reserved.

I dedicate this dissertation to my parents, Darrin and Nicole, my grandparents, Victor, Dianne, Jerry, and Helen, and my fiancè, Wencel. I wouldn't have made it this far without your love and support.

Acknowledgments

First and foremost, I would like to express my sincerest gratitude to my advisor, Dr. Nina Fefferman, for her unending support, encouragement, and guidance. Not only is she a talented mathematician and researcher, but also one of the kindest people I know. She will continue to be a role model in both my professional and personal life.

I want to thank Dr. Judy Day for her support and guidance as I began my research. I am also extremely appreciative of my other collaborators, Dr. Patricia Bamwine, Dr. Kristen Rappazzo, and Dr. Tom Luben. I want to thank the remaining members of my committee, Dr. Chien-fei Chen and Dr. Olivia Prosper for their time and helpful feedback.

I am thankful to the math department for giving me the opportunity to pursue my PhD and for all the help and support I received from the faculty and staff. I especially want to thank my classmates, labmates, and cohort-mates for their friendships and community. Graduate school would not have been the same without them.

I am incredibly grateful to my parents, Nicole and Darrin, my siblings, Tommy, Mary, Joey, and Michael, and my grandparents, Jerry, Helen, Victor and Dianne, for always believing in me. I also want to thank my dog, Gauss, for his unconditional love and companionship. Lastly, I want to thank my fiancé, Wencel, who was by my side supporting and encouraging me throughout this whole process.

Abstract

Mathematical modeling is an important tool for analyzing and understanding the world around us. Throughout history, it has been instrumental in advancements made in the sciences. Recently, mathematical modeling has been used to gain insight in the social sciences, specifically social issues. Thus, the overall goal for this dissertation is to build up the mathematical modeling toolkit for addressing social issues, especially by applying and developing techniques in network theory. We accomplish this by looking at three social issues facing our communities.

In Chapter 2, we use flow networks to develop an alternative method for calculating and analyzing the basic reproductive number for infectious disease outbreaks, which is the expected number of secondary infections produced by a typical infectious person in a susceptible population. We developed this method to be more accessible to non-mathematical audiences, as the basic reproductive number is important for public health officials and the general population in understanding infection risk and responding to outbreaks. We show our method is equivalent to traditional methods and provide instructions on its implementation.

Chapter 3 applies immuno-epidemiological modeling techniques to study violence spread through exposure. Recently, there has been a push to understand violence as a public health issue. We expand this analogy between violence and infectious diseases to formulate a susceptible-exposed-infectious model. We then provide stability and equilibrium analysis and run example numerical simulations to show that the insights gained from a mathematical

model might help identify effective interventions.

The fourth chapter uses network modeling to explore how public transportation can help connect food desert residents to grocery stores. According to the USDA, food deserts are census tracts that experience high poverty rates and limited grocery store access. We select five cities and formulate network models where the food deserts and grocery stores are the nodes and transit lines are the edges. We analyze these networks through centrality measures and provide policy suggestions to improve public transit use in increasing food access.

In total, this work presents examples of how novel mathematical models (especially using network theory) can help address societal problems and effect meaningful change.

Table of Contents

1	Introduction	1
2	A Network Theoretic Method for the Basic Reproductive Number	7
2.1	Introduction	9
2.2	Max-Flow Min-Cut Method	12
2.2.1	Notation	13
2.2.2	Equivalence Between Methods	14
2.3	Algorithm and Examples	16
2.4	Discussion	23
2.5	Additional Examples	25
2.6	Conclusion	32
3	Linking Immuno-Epidemiology Principles to Violence	34
3.1	Introduction	36
3.2	Methods	39
3.2.1	Analogy	39
3.2.2	Model Description	40
3.2.3	Model Parameters	43
3.2.4	Equilibrium and Stability Analysis	45
3.2.5	Numerical Simulations	49
3.3	Results	50
3.3.1	Equilibrium and Stability Analysis	50
3.3.2	Numerical Simulations	58

3.4	Discussion	61
3.4.1	Equilibrium and Stability Analysis	61
3.4.2	Numerical Simulations	62
3.4.3	Public Health Policy	64
3.4.4	Limitations of the Model	65
3.5	Basic Reproductive Number	66
3.5.1	Contributive Number	67
3.6	Conclusion	72
4	Connecting People to Food: A Network Approach to Alleviating Food Deserts	74
4.1	Introduction	76
4.2	Methods	78
4.2.1	Model Formation	78
4.2.2	Data	79
4.2.3	Centrality Measures	82
4.2.4	Sensitivity of Representative Addresses	84
4.3	Results	86
4.3.1	Degree	86
4.3.2	Closeness	90
4.4	Discussion	93
4.4.1	Limitations	97
4.5	Conclusion	98
5	Conclusion	100
	Bibliography	107
	Appendices	118
A	Appendix for Chapter 4	119
A.1	Sensitivity Analysis of the Representative Addresses	141
A.2	Path Length Analysis for a Theoretical Food Desert Network	147

List of Tables

3.1	Parameters	46
3.2	Equilibrium Points	51
3.3	Nullclines	51
4.1	SIC Codes	81
4.2	Wilcoxon Signed-Rank Tests for Degree in Brown Deer	89
4.3	Wilcoxon Signed-Rank Tests for Degree in Lawrence	89
4.4	Wilcoxon Signed-Rank Tests for Closeness in Brown Deer	92
4.5	Wilcoxon Signed-Rank Tests for Closeness in Lawrence	92
1	Representative Addresses for Brown Deer	122
2	Representative Addresses for Lawrence	122
3	Representative Addresses for Albuquerque	123
4	Representative Addresses for Charlotte	124
5	Representative Addresses for Raleigh	125
6	Overlap Centrality Results	126
7	Degree Results for Brown Deer	131
8	Degree Results for Lawrence	131
9	Degree Results for Albuquerque	132
10	Degree Results for Charlotte	133
11	Degree Results for Raleigh	134
12	Closeness Results for Brown Deer	134
13	Closeness Results for Lawrence	135
14	Closeness Results for Albuquerque	135

15	Closeness Results for Charlotte	136
16	Closeness Results for Raleigh	137
17	Closeness Results for Brown Deer with Unverified Grocery Stores	137
18	Closeness Results for Lawrence with Unverified Grocery Stores	138
19	Closeness Results for Albuquerque with Unverified Grocery Stores	138
20	Closeness Results for Charlotte with Unverified Grocery Stores	139
21	Closeness Results for Raleigh with Unverified Grocery Stores	140
22	Representative Addresses for Brown Deer Selected with the Random Method	142
23	Representative Addresses for Brown Deer Selected with the Center Method .	142
24	Representative Addresses for Lawrence Selected with the Random Method .	143
25	Representative Addresses for Lawrence Selected with the Center Method . .	143
26	Degree Results for Brown Deer with Representative Addresses Selected Using the Random Method	143
27	Degree Results for Brown Deer with Representative Addresses Selected Using the Center Method	144
28	Degree Results for Lawrence with Representative Addresses Selected Using the Random Method	144
29	Degree Results for Lawrence with Representative Addresses Selected Using the Center Method	144
30	Closeness Results for Brown Deer with Representative Addresses Selected Using the Random Method	145
31	Closeness Results for Brown Deer with Representative Addresses Selected Using the Center Method	145
32	Closeness Results for Lawrence with Representative Addresses Selected Using the Random Method	145
33	Closeness Results for Lawrence with Representative Addresses Selected Using the Center Method	146
34	Path Length Analysis for Original Network	154
35	Path Length Analysis for Network with Bus Stops Added Randomly	155
36	Path Length Analysis for Network with Bus Stops Added Manually	156

37	Path Length Analysis for Network with Edges Removed	157
----	---	-----

List of Figures

2.1	Flow Network Example	11
3.1	Compartment Diagram for One Exposed Class Model	44
3.2	Compartment Diagram for Three Exposed Class Model	47
3.3	Nullcline Plot	52
3.4	Phase Portrait	54
3.5	Numerical Simulations	59
3.6	Unweighted Flow Network	69
3.7	Weighted Flow Network	69
3.8	Flow Network Evaluated at VFE	70
3.9	Reduced Flow Network	70
3.10	Cut Sets for Flow Network	71
4.1	Full Network for Brown Deer.	87
4.2	Degree Results	88
4.3	Closeness Results	91
1	Full Network for Lawrence	127
2	Full Network for Albuquerque	128
3	Full Network for Charlotte	129
4	Full Network for Raleigh	130
5	Theoretical Food Desert Network	149
6	Network with 10 Bus Stops Added Randomly	151
7	Network with 5 Bus Stops Added Manually	152

8	Network with 5 Random Edges Removed	153
---	---	-----

Chapter 1

Introduction

One of humanity's greatest achievements is the eradication of smallpox in 1980. Smallpox was a terrible disease that had a fatality rate anywhere from 20 – 60 percent (with the fatality rate in infants being much higher). Those who managed to survive the infection were often left disfigured and disabled [71]. The goal of managing smallpox had been inspiring researchers and scientists long before the eradication campaign in the 1960s, 1970s and 1980s [71, 85, 11]. The development of vaccines themselves started with cowpox inoculations against smallpox in 1796 [71]. In addition to advancements in medicine and biology, smallpox research also inspired advancements in mathematics. The first ever mathematical model of an infectious disease was of smallpox by Daniel Bernoulli in 1760 [11]. Bernoulli's work on smallpox is generally thought to be the start of the field of mathematical epidemiology [13] (to which some of the work in this dissertation contributes). While this was the first instance of mathematical modeling applied to epidemiology, this was by no means the first use of mathematical modeling to describe the real world. In fact, the use of mathematical models is as old as mathematics itself [78] and new research and advancements continue to this day. We now have numerous modeling techniques, such as dynamical systems, network theory, game theory, agent-based models, partial differential equations models, and stochastic models, to address problems across all fields of science. Recently, mathematical modeling has also started being widely used in the social sciences as a tool to tackle social issues. This idea of applying mathematical modeling to social issues, more specifically to issues that face

communities, is an underlying theme of this dissertation.

This dissertation has three distinct chapters, but the theme throughout is using mathematical modeling to understand and aid in alleviating societal issues that communities face. Broadly, the three community issues explored here are disease spread, violence, and food access. To gain insight into these topics mathematically, the primary tools we use are network theory and dynamical systems. Overall, this dissertation is motivated by trying to make an impact, using mathematics, on problems that people are struggling with in the real world. While the issues we look at present unique and interesting mathematical challenges, they are interdisciplinary in nature and thus their results are also of interest to those in other fields, such as public health and policy, violence prevention, food security, and epidemiology. Due to this interdisciplinary nature, one of the goals for this dissertation is for the results to be practically useful, not only for those in mathematics, but also for those not in mathematics. Specifically, for Chapters 3 and 4 one of the goals of the model analysis is to suggest concrete policy steps that can be taken to aid in resolving the respective issues of each chapter.

While the social application in each chapter is different, there is overlap in the mathematical tools. All three chapters have a network theory component. Network theory, originally referred to as graph theory, is the study of networks (graphs). A graph is defined as ordered pair, (V, E) , where V is the set of nodes and E is the set of edges. Every element in E connects two nodes, or a node to itself [40]. Graph theory developed out of the Königsberg Bridge Problem, which questioned if it was possible to find a tour of Königsberg (a city in Prussia) that crossed each of the city's bridges only once and returned the tourist to their original starting point [40]. The solution to this problem was published in 1736 by mathematician Leonhard Euler, proving that it was impossible. This is generally considered the first result in graph theory. From this first puzzle, graph theory (and later, network theory) expanded into a field of mathematics itself and as a fundamental method and tool in other fields, including chemistry [7, 9], physics [32, 51], computer science [69, 70], and epidemiology [29, 35] to name a few.

The other main modeling technique used in this dissertation is dynamical systems, more specifically epidemiological compartmental models. We see this modeling technique most heavily relied on in Chapter 3, but an understanding of it is also needed in Chapter 2. In general, epidemiological compartmental models are ordinary differential equations (ODEs) that capture disease spread on a community level. This is done by dividing a population into various compartments or classes, the most common being susceptible (S), infected (I), and recovered (R). This is referred to as a *SIR* model [82]. There are many variants of this model depending on the classes included or excluded. An exposed or vaccinated class are other common compartments to include. Such models can also be adjusted for other epidemiological dynamics, such as vector-host dynamics and multi-strain diseases. Compartmental modeling of infectious diseases was developed in the early twentieth century, first by Ronald Ross and Hilda Hudson [74, 75, 48] and then expanded on by Anderson Gray McKendrick and William Ogilvy Kermack [50]. Advances in compartmental modeling and analyzing of infectious diseases continued through the rest of the twentieth century with the works of people like Odo Diekmann, Hans Heesterbeek, Hans Metz, and Klaus Dietz [27, 28] and have continued up to the present with research from people like Pauline van den Driessche, James Watmough, Fred Brauer, and Hans Heesterbeek [14, 91, 46]. Next we will give a brief summary of each chapter, but a more thorough introduction for each project is provided in their respective chapters.

In the second chapter we explore a new way to calculate and analyze the basic reproductive number, $\mathcal{R}_0$. Epidemiologically, $\mathcal{R}_0$ is defined as the expected number of secondary infections produced by a typical infectious person in a completely susceptible population [27], while mathematically $\mathcal{R}_0$ is defined as the threshold parameter for invasion of an infectious pathogen into a completely susceptible population [14]. The basic reproductive number is one of the most important tools in epidemiology for understanding outbreaks and emerging epidemics of infectious disease. Thus, finding a new and more accessible way to calculate and understand it can help communities better respond when an outbreak occurs. This is important as public health leaders may not always have access to mathematical scientists to assist with the calculation and interpretations of $\mathcal{R}_0$ in real time, which could cause a

delay in their response. So, the question that motivates this chapter is: is it possible to develop a new algorithm that is both equivalent to the current state of the art method, the Next Generation Matrix method [27, 91], and more accessible to broad audience? To answer this question we approach it from a network theory perspective, more specifically using flow networks. We develop an algorithm to convert traditional compartmental ordinary differential equation models to flow networks and then use the Max-Flow Min-Cut Theorem to find the basic reproductive number. We provide step-by-step instructions on how to implement our method and show it is equivalent to the Next Generation Matrix method. Then we demonstrate its application by computing $\mathcal{R}_0$ for both general models and specific examples from the literature. Further, we discuss the novel insights a network based method provides that a traditional compartment model cannot by analyzing the cut sets themselves.

The next chapter of the dissertation expands on the idea of treating violence as an infectious disease. Every day approximately 4400 people lose their lives to intentional acts of violence worldwide [19]. As viewing violence as a public health issue is gaining popularity [81], so is employing traditional public health solutions and tools. A fundamental tool used to understand transmission and spread dynamics of infectious disease is mathematical epidemiological compartment models. Thus, our research question is: by building upon the emerging framework of treating violence as an infectious disease and viewing it as a public health issue, can we use traditional epidemic models to learn about the dynamics of violence spread and come up with viable solutions for how to reduce the effects of it in our communities? In this chapter, we expand and formalize the analogy between violence and infectious diseases so we can develop a community level model to track how violence spreads through exposure. After formulating the model, we analyze it through stability and equilibrium analysis and by generating some example numerical simulations. These analysis show potential insights that can be gained by employing a model like this to help combat violence. We also develop a new way to think about a basic reproductive number-type concept that would apply to violence and other epidemiologically-based social models.

Finally, in Chapter 4 we look at the issues surrounding food access in our communities, specifically food deserts. Food deserts are census tracts that are both low income and low access, where low income is defined by the United States Department of Agriculture (USDA) to have a poverty rate of 20 percent or greater, or a median family income at or below 80 percent of the statewide or metropolitan area median family income, and low access is defined to be at least 500 persons and/or at least 33 percent of the population living more than 1 mile from a supermarket or large grocery store [68]. Food deserts exist in all fifty states and are a major public health concern for the communities they affect, as those living in them lack adequate access to nutritious foods. Since city leaders do not have full control of where grocery stores open, we want to explore ways that city leaders can help their food desert residents access grocery stores. Thus, our research question is: can public transportation networks be used to help alleviate food deserts? We pick five cities across the country to study by formulating a network out of the food deserts, grocery stores, and public transit, and analyze it using centrality measures. We then provide policy recommendations for city leaders that are based on our analysis for how to utilize their public transit systems to alleviate food deserts.

All three of the projects presented here help to expand the mathematical toolkit available to tackle societal-based community issues, especially by applying techniques from network theory. Network theory is a fundamental tool to each of the three chapters. In Chapter 2 we use a specific type of network, a flow network, to make an advancement in epidemiology by developing a new way to calculate and view $\mathcal{R}_0$ and this is further expanded in Chapter 3 by using this flow networks based method to apply the idea of a basic reproductive number outside of its traditional epidemiological purpose. Chapter 4 uses a more classical network model formulation and analysis to gain understanding of an important social issue, food access. These chapters help lay the groundwork to expand the use of mathematical modeling to other community level issues where modeling may not have been considered before. While this dissertation focuses explicitly on disease spread, violence, and food access the ideas developed here can easily be adjusted and applied to other important community level problems, such as misinformation spread and employment access. In addition, mathematical

models serve as decision support tools for policy makers. Models can provide quantitatively based suggestions about the most efficient ways to address a problem. Models also offer a cost-effective way to study an issue. By using a model, community leaders (or their analysts) can run simulations to test the outcome of different options before implementing one in their community. Therefore, the work in this dissertation helps expand both the tools available in applied mathematics and the use of modeling to aid in the policy responses to issues affecting our communities.

Chapter 2

A Network Theoretic Method for the Basic Reproductive Number

Disclosure statement: Anna Sisk is the first and corresponding author on the manuscript based on this chapter. Her co-author, Nina Fefferman (advisor), directed the work. Sisk and Fefferman worked together to develop the concept, show the equivalence, formalize the method, and execute the examples. Both authors reviewed, revised, wrote, and approved the manuscript.

Publication information: In review

Abstract

The COVID-19 pandemic has highlighted the epidemiological importance of the basic reproductive number, $\mathcal{R}_0$, as a tool for both public and scientific understanding of infection risk. While estimates for the realized rate of increase, $\mathcal{R}_{eff}$, can be estimated by curve-fitting to incidence data, available methods for calculating $\mathcal{R}_0$ as a means of understanding and predicting overall dynamics and future outbreak behavior are complicated, requiring a mathematical background to learn and implement appropriately. This can act as a barrier to understanding for public health teams that do not already work with mathematical scientists and cause delays in a real-time response. In this chapter, we present a new method for calculating and analyzing the basic reproductive number that is far more accessible to the broader scientific community. This network-based approach drastically reduces the complexity and mathematical background needed to understand the methodology and perform the calculation. We show that this more accessible method is equivalent to the traditional Next Generation Matrix method for the calculation of $\mathcal{R}_0$, and thus a valid alternative. Further we provide step-by-step instructions for this new method and examples to illustrate how to apply it to epidemic models. Finally, we discuss the novel insights a network-based approach provides when analyzing $\mathcal{R}_0$ and how it can be applied to models and situations where traditional methods fail.

2.1 Introduction

One of the most influential developments in epidemiology and public health is the characterization of the basic reproductive number, often referred to as $\mathcal{R}_0$. While originating in the fields of demography and epidemiology [46], the basic reproductive number has become an important aspect of mathematical biology and is now mathematically defined and quantifiable [28, 27, 91]. Epidemiologically, $\mathcal{R}_0$ is defined as the expected number of secondary infections produced by a typical infectious person in a completely susceptible population [27], while mathematically $\mathcal{R}_0$ is defined as the threshold parameter for invasion of an infectious pathogen into a completely susceptible population [14]. Calculating $\mathcal{R}_0$ offers important information in responding to an outbreak as its value can help determine if a disease can invade a population. If $\mathcal{R}_0 > 1$ the infectious individuals will replace themselves and increase with each generation, but if $\mathcal{R}_0 < 1$, infectious individuals will fail to replace themselves and thus the disease will die out. In addition to determining whether a disease will successfully take hold in a population, the value of $\mathcal{R}_0$ gives an idea of how quickly disease will initially spread. A larger value for $\mathcal{R}_0$ implies the disease will spread quicker, like measles, which some studies show can have a basic reproductive number greater than 50 [44], while a smaller $\mathcal{R}_0$ points towards a slower spread, as seen in Ebola which has reported $\mathcal{R}_0$ values of less than 3 [5]. Thus, mathematical analysis of $\mathcal{R}_0$ is an important step in both describing and predicting a disease outbreak and in formulating a response.

Analyzing $\mathcal{R}_0$ can also provide critical insight for how to decrease its value and therefore decrease the spread of infection. This type of analysis is very commonly seen in the mathematical biology and epidemiology literature [90, 55, 61, 56, 83]. Depending on the model type there are different ways to calculate $\mathcal{R}_0$. The most basic way is through visual inspection. This method will usually only work for simple Susceptible-Infected-Recovered (SIR)-type models by determining what makes the infectious equation, or $\dot{I}$, positive and therefore the infectious class increasing. This method leads to the definition of the basic reproductive number as: (duration of infectious period)*(the rate of secondary infections). While this is a quick and easy way to determine $\mathcal{R}_0$ it only works for simplistic models. If we

want to look at more sophisticated models, such as ones with exposed classes, multi-strains, or vector-host dynamics, we need a more sophisticated method. The Next Generation Matrix (NGM) method, first introduced in [27] and built upon by [91], allows for $\mathcal{R}_0$ to be calculated for these more complicated models. This method, however, is quite mathematically technical and not easily accessible for broad scientific practice [15, 61]. While these two methods are the most common to find the basic reproductive number, there are additional methods discussed in [54].

One approach to understanding and calculating $\mathcal{R}_0$ discussed in [54] is through graph theory. Graph, or network, theory is a branch of mathematics that can offer a different perspective on disease dynamics and modeling [72]. A graph, also referred to as a network, is comprised of two sets, vertices and edges. Vertices, also called nodes, are connected to each other by edges. This map of connections of vertices by edges is called a graph or network. Networks are often represented visually by a digraph. Figure 2.1 shows an example of this. Network models offer insight to disease transmission dynamics that may be overlooked when using other model types. They can be especially helpful in explaining mathematical models and results to a non-mathematical audience, as networks can be more intuitive than other model types (e.g. ordinary or partial differential equations). Networks have been used to study outbreaks of infectious diseases, such as sexually transmitted diseases [29], COVID-19 [35], and to learn about general epidemic dynamics [49]. Thus, graph theoretic approaches have already been explored as a way to learn about the basic reproductive number. One way this is done is in conjunction with the NGM [23]: a digraph is formulated from components of the NGM and a reduction algorithm is used to calculate $\mathcal{R}_0$. Network theory is also used to analyze $\mathcal{R}_0$, such as in [52] which studies the effect that edge removal has on the size of $\mathcal{R}_0$.

Thus far we have discussed two of the main ways to model infectious diseases. First, is using compartmental ordinary differential equation (ODE) models. This approach tracks how population segments move through health states. This is a community level approach to modeling. To model infectious disease spread on the individual level network models can be used. Networks model how individuals infect each other. We are proposing a third method

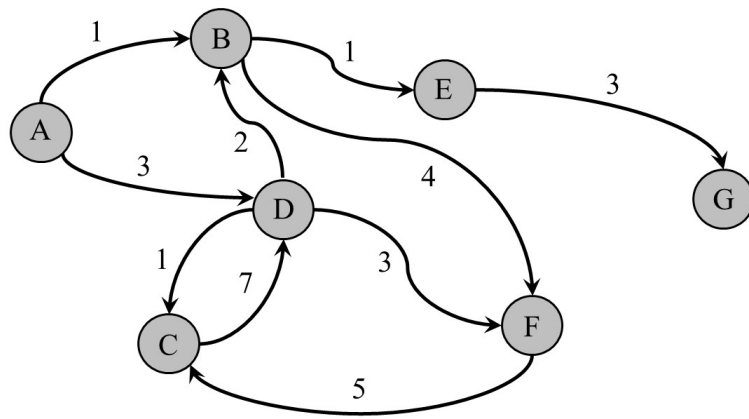


Figure 2.1: Flow Network Example. Example of a weighted digraph where the circles with the letters inside are the nodes, the arrows are the edges, and the numbers above the edges are the edge weights.

to look at infectious disease spread: modeling how infection itself flows through a network of population segments. Using this third method we develop a network theoretic approach to calculating and analyzing $\mathcal{R}_0$. To do this we use a specific type of network model – a flow network. A flow network model (sometimes called a transportation network) is a directed graph in which the nodes and edges have capacities to handle “flow” through them. Flow networks also have a source node and a sink node. A source node is the point of origin for all flow into the network and the sink node is where the flow terminates. No flow within the network may enter the source or exit the sink. For an epidemiological model, we will consider the disease to be the flow and the infectious class to be the source. (Note the slight difference in framing from standard ODE SIR models, in which individuals are the objects moving through the system [14].) To use a flow network to compute the basic reproductive number for a disease, we therefore must first convert a traditional ODE compartmental model into a flow network. To do this, we consider each class as a node in the network and study the ‘flow’ of disease through the nodes. In this way, we develop an alternative approach to calculating the basic reproductive number and show this new method is equivalent to the NGM method for finding $\mathcal{R}_0$. We also present an algorithm for converting compartmental ODE models to flow networks and give instructive examples of how to use this method to calculate the basic reproductive number.

2.2 Max-Flow Min-Cut Method

In order to calculate the basic reproductive number with a flow network, we need to determine the maximum amount of flow, or disease, that can be sent from the source to the sink. In an epidemiological context, this will be the maximum amount of disease generated from one infectious individual at the start of an outbreak, the source, that will end up in the sink, and therefore be available to the infectious class (the source) in the next generation. In other words, the amount of flow that ends up in the sink represents the number of infections there will be in the next generation. To calculate this we need to solve the max flow problem, which is a fundamental problem in network theory [20, 41]. The technique for calculating the

max flow is finding the minimum cut, which by the Max-Flow Min-Cut Theorem, is equal to the max flow.

Theorem 2.1. *For any flow network the value of the maximum flow from the source to the sink is equal to the weight of the minimum cut over all cuts that separate the source and sink. [37]*

Thus, once we determine the max flow, we can analyze it in a similar way to the traditional methods for finding the basic reproductive number. If the flow is positive, then the disease will continue to spread and, if it is zero or negative, then the disease dies out. Hence we will call this new method for calculating $\mathcal{R}_0$ the Max-Flow Min-Cut (MFMC) method. Note that the main difference between modeling an outbreak as a network versus a compartmental ODE model is the network model tracks the movement of infection in a population, while the ODE model tracks the movement of people through health stages.

2.2.1 Notation

Before we can show that the MFMC method is equivalent to the NGM method we must first define notation. We start with some general network notation. Let $N = (V, E)$ be a directed network where V is the set of all the vertices and $E = \{(u, v) | \forall u, v \in V\}$ is the set of all the directed edges (also called ‘arcs’). Further, let $w(u, v)$ be the weight of the (u, v) edge. Specifically for flow networks, we denote the source as $s \in V$ and the sink as $t \in V$. Now we can develop notation for our application. Note that the edges originating from s represent new infections. Denote this set of edges as Φ . Then denote the rest of the edges as Ψ . Thus, for a network, N , formulated from a compartmental ODE model $E = \Phi + \Psi$. Since we have notation for the edges, let’s define notation for the edge weights. Let $w(\Phi)$ represent the total weight of all the edges in Φ and $w(\Phi_{u,v})$ represent the weight of the specific edge, $(u, v) \in \Phi$. We will use the same notation for the edges in Ψ .

For the next generation matrix we employ the notation from [14] and [91]. Let x represent the disease compartments and let y represent the infection-free compartments. For this calculation, we are only interested in the infected compartments. Thus, we can write the

infected compartments of the model as follows,

$$x'_i = \mathcal{F}_i(x, y) - \mathcal{V}_i(x, y), i = 1, \dots, n$$

where $\mathcal{F}_i$ is the rate of secondary infections that increase the i^{th} infected compartment and $\mathcal{V}_i = \mathcal{V}^- - \mathcal{V}^+$ is the rate of disease removal ($\mathcal{V}^-$) and progression ($\mathcal{V}^+$) that decrease the i^{th} compartment. Further [14] and [91] also explain what the entries of the the next generation matrix represent: the (i, j) entries of V^{-1} are the expected times an individual initially introduced into the infection compartment j spends in infection compartment i , the (i, j) entries of F are the rate secondary infections are produced in compartment i by an index case in compartment j , and so the (i, j) entries of FV^{-1} are the expected number of secondary infections in compartment i produced by individuals initially in compartment j .

2.2.2 Equivalence Between Methods

Assume we are modeling a closed system with a well-mixed population at the beginning of an outbreak. Thus, we want to show that the MFMC method and the NGM method provide equivalent calculations for the value of $\mathcal{R}_0$. First, let's consider how the next generation matrix (FV^{-1}) and the network (N) model secondary infections and their spread. In the NGM method, $\mathcal{F}$ is the rate at which secondary infections increase the infection compartments. In the MFMC method, Φ is the set of edges that represent the rates at which secondary infections spread to the rest of the nodes. Thus, $\mathcal{F}$ and Φ are qualitatively equivalent. The same is true for $\mathcal{V}$ and Ψ as they both represent the rest of the movement in the infection compartments.

In addition to both methods representing the same qualitative ideas, they are also quantitatively equivalent with some regrouping. Note,

$$\begin{aligned} w(\Phi) &= \mathcal{F} + \mathcal{V}_I^- \\ w(\Psi) &= \mathcal{V}_{i^*}^+ + \sum_{j \neq I} \mathcal{V}_{j \neq I}^- \end{aligned}$$

where $\mathcal{V}_I^-$ represents removal terms from the I class, $\sum \mathcal{V}_{j \neq I}^-$ represents the rest of the removal terms, and $\mathcal{V}_{i^*}^+$ represents all of the progression terms except those that allow for t to feed directly into s . Note, the exclusion of these terms does not cause a difference in $\mathcal{R}_0$, instead they are essentially functioning as lag terms (which do not affect the contribution of new infections towards the next generation). These lag terms also disappear while computing FV^{-1} , see Section 2.5 for more details. Thus, $\mathcal{F}, \mathcal{V}$ and $\Phi, \Psi, w(\Phi), w(\Psi)$ are equivalent in both mathematical expression and interpretation.

Next, the NGM method computes $\frac{\partial \mathcal{F}}{\partial x}|_{(0, y_0)} = F$ and $\frac{\partial \mathcal{V}}{\partial x}|_{(0, y_0)} = V$ to consider the behavior of the system at the beginning of an outbreak. The MFMC method also considers the system behaviors near the disease free equilibrium (DFE), but does so by evaluating $N|_{(S^* \approx P, I^* > 0)}$, where P is the total population and I^* is small. While these methods for analyzing behaviors near the DFE are different, they are equivalent.

Since we have seen that FV^{-1} and N have equivalent methods for modeling the amount of secondary infections per generation that occur at the beginning of an outbreak, this implies if there is a difference between these two methods it must occur in calculating $\mathcal{R}_0$. In the NGM method $\mathcal{R}_0 = \rho(FV^{-1})$, where ρ is the spectral radius. The MFMC method calculates $\mathcal{R}_0$ by finding the minimum cut, which by the Max Flow Min Cut Theorem, is the maximum flow. Therefore each technique calculates $\mathcal{R}_0$ by finding dominant influence in the long term behavior of the system or network. Since both techniques model secondary infections equivalently, calculating $\mathcal{R}_0$ by finding the dominant influence of the system by the spectral radius is equivalent to finding the dominant influence in a network by maximum flow.

From these arguments, we see that the Next Generation Matrix method and the Max Flow Min Cut method are equivalent in both their understanding and calculation of secondary infections at the beginning of an outbreak.

2.3 Algorithm and Examples

To get a better understanding of both techniques and their equivalence, we present each algorithm for execution and use both to work through two examples. Consider a closed system, compartmental epidemic model. Assume a disease well-described by a basic *SIR* model [14],

$$\begin{aligned}\dot{S} &= -\beta SI \\ \dot{I} &= \beta SI - \gamma I \\ \dot{R} &= \gamma I,\end{aligned}\tag{2.1}$$

where β is the contact rate and probability of transmission of infection between the susceptible and infectious and γ is the removal rate for the infectious class (either through recovery or death). Further, let N be the total population (N will be the total population for all the following examples in this chapter as well). Due to the simplicity of the model we can compute $\mathcal{R}_0$ by inspection. First, we evaluate the model near the DFE ($S^* \approx N, I^* > 0$, but small) and then solve for $\dot{I} > 0$. Thus, the basic reproductive number is $\beta S^*/\gamma$. For instructive purposes, we now present the respective step-by-step algorithms for finding $\mathcal{R}_0$ using both the NGM and MFMC methods.

Next Generation Matrix Method:

- a) Divide the classes into two compartments: disease, x and non-disease, y . We are only interested in the disease compartment.

- $x = I$
- $y = S, R$

- b) Determine which terms from the equations for the classes in the disease compartment represent new infections in the disease compartment, $\mathcal{F}$, and which represent movement among, $\mathcal{V}^+$, or out of the disease compartment, $\mathcal{V}^-$.

- $\mathcal{F} = \beta SI$

- $\mathcal{V} = \gamma I$

c) Take the derivative of $\mathcal{F}$ and $\mathcal{V}$ with respect to the state variables of the disease compartment, x . We call the resulting matrices F and V .

- $\frac{\partial \mathcal{F}}{\partial x} = \beta S = F$
- $\frac{\partial \mathcal{V}}{\partial x} = \gamma = V$

d) Evaluate F and V at the DFE.

- DFE: $(S^*, 0, 0)$
- $F|_{(S^*, 0, 0)} = \beta S^*$
- $V|_{(S^*, 0, 0)} = \gamma$

e) Compute V^{-1} and then FV^{-1} .

- $V^{-1} = \frac{1}{\gamma}$
- $FV^{-1} = \frac{\beta S^*}{\gamma}$

f) Determine $\rho(FV^{-1}) = \mathcal{R}_0$. This, along with its eigenvector represents the distribution of infected individuals that produces the greatest number of secondary infections per generation.

- $\mathcal{R}_0 = \frac{\beta S^*}{\gamma}$

Max Flow Min Cut Method:

1. Divide the compartments into two groups: compartments through which infection can flow and compartments through which it cannot. We are only interested in the compartments that act as conduits for infection propagation. These compartments will become the nodes in our network.

- Infection Flows Through: I, S
- Infection Does Not Flow Through: R

2. Identify the infectious class(es). This is the compartment(s) that can transmit the infection. This compartment will be the source node, s .

- $s = I$

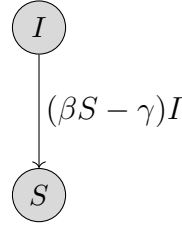
3. Identify the compartment(s) that connects, or feeds directly, into the infectious class(es). This compartment will be the sink node, t .

- $t = S$

4. The remaining compartments will be the intermediary nodes in the network. All of the nodes in the network are connected by edges. The set of edges that leave the source node and represent the new infections in the community make up Φ . While the rest of the edges that represent the movement of the infection are in the set Ψ , including the edges that have repeated weights due to conservation of flow. $w(\Phi), w(\Psi)$ come directly from the terms in the compartmental model.

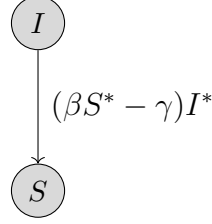
- $\Phi = \{(I, S)\}, w(\Phi) = (\beta S - \gamma)I$

- $\Psi = \emptyset$



5. The compartmental ODE model is now formulated into a network model. Now we evaluate the edges near the DFE. In other words, we let the infectious class be small, but positive, and the susceptible class be approximately the entire population: $(S^* \approx N, I^* > 0, \text{ but small})$.

- Evaluate near DFE, $(S^*, I^*, 0)$:



6. Find the cut sets, which are the sets of edges that, when cut, separate s from t . The cut sets represents the ways the disease can spread through a community to cause secondary infections and the cuts show how the spread can be stopped. The capacity of the cut represent the amount of disease/infection available for secondary infections.

- Cut set: $\{(I, S)\}$

7. Determine the cut set with the minimum edge weight. By the Max Flow Min Cut Theorem, we know that is also the max flow.

- Minimum cut weight = Maximum flow: $(\beta S^* - \gamma)I^*$

8. Solve the equation for the epidemic growth condition desired. For example, for an outbreak to occur (i.e., positive growth), the max flow must be positive, so we can solve the inequality for $\mathcal{R}_0 > 1$.

- Max flow $> 0 \implies (\beta S^* - \gamma)I^* > 0 \implies \mathcal{R}_0 = \frac{\beta S^*}{\gamma} > 1$

Now that we have defined the algorithms and worked through a concrete example implementing both methods, we can gain better intuition into the MFMC method by examining the step-wise equivalence between the methods. First, we note that both techniques start off similar by considering where there is infection in the model. A subtle difference is that while the NGM method looks at where the infected (not just infectious) people are in the compartmental model, the MFMC method looks at which compartments house infection and where it can spread from there. Thus, 1. and a) are equivalent. Next, both techniques work to identify the way each of the terms in the compartment model move, either infected people (NGM method) or the infection (MFMC method). Thus, b) and 2.-4. are equivalent. Specifically, we note that in the MFMC method, the edges that leave the

source play the same role that $\mathcal{F}$ does in the NGM method, and the rest of the edges in the network represent the same type of movement as $\mathcal{V}$. Since these techniques are used to look at the early stages of an outbreak, both techniques focus on how the system behaves near the DFE. The NGM method does this by linearizing about the DFE (c)-d)), while the MFMC method does this by analyzing what the system does when it is just off the DFE, (5.). This is similar to the more basic technique for solving for $\mathcal{R}_0$ by finding which parameter combinations make $\dot{I} > 0$ near the DFE. Finally, the interpretations of the entries FV^{-1} and its eigenvalues/vectors (including $\rho(FV^{-1})$) in e)-f) and the interpretations of the cut sets, edge capacities, and max flow seen in 6.-7. consider the same idea from different perspectives. The NGM method considers how infected and infectious individuals move through the classes and produce secondary infections, while the MFMC method considers how the infection moves through the nodes and how many infections there will be to contribute towards new infections in the next generation.

Keeping this step-wise equivalence in mind, let us consider the example. Since the NGM method is an established technique, we will focus on how the MFMC algorithm was applied. First, we note that the only class that is not explicitly included in the network is the R class, as disease cannot flow into it. However, the movement of infectious individuals into the R class does have an effect on the amount of infection in the network. Since infectious individuals are moved to the R class at rate γ , that is also the rate at which infection is removed from the network. Thus, since the network is modeling a community, not an individual, we need to take γ into consideration when determining the overall amount of infection that can flow into the community. Hence, the amount of infection that can flow from the infectious node to the rest of the network is the amount of effective contacts minus the amount that has been removed/recovered, which can be thought of as the net flow of new disease. So, the weight of the only edge in the network is $\beta S^* I^* - \gamma I^*$, which is equal to $\mathcal{F} - \mathcal{V}$.

Next, let us consider a model with slightly more complexity by adding in an exposed class. The *SEIR* model [14] is,

$$\begin{aligned}
\dot{S} &= -\beta SI \\
\dot{E} &= \beta SI - \kappa E \\
\dot{I} &= \kappa E - \gamma I \\
\dot{R} &= \gamma I,
\end{aligned} \tag{2.2}$$

where β is the contact rate, κ is the rate of disease progression, and γ is the removal rate. Unlike the first example, determining the basic reproductive number by inspection fails due to the inclusion of the exposed class. So, for more complicated models, a method like the NGM or MFMC is most often used to analytically determine $\mathcal{R}_0$. The calculations below follow the same steps as defined above in each of the algorithms.

Next Generation Matrix Method:

$$\text{a) } x = I, E$$

$$y = S, R$$

$$\text{b) } \mathcal{F} = \begin{pmatrix} \beta SI \\ 0 \end{pmatrix}$$

$$\mathcal{V} = \begin{pmatrix} \kappa E \\ \gamma I - \kappa E \end{pmatrix}$$

$$\text{c) } \frac{\partial \mathcal{F}}{\partial x} = \begin{pmatrix} 0 & \beta S \\ 0 & 0 \end{pmatrix} = F$$

$$\frac{\partial \mathcal{V}}{\partial x} = \begin{pmatrix} \kappa & 0 \\ -\kappa & \gamma \end{pmatrix} = V$$

$$\text{d) DFE: } (S^*, 0, 0, 0)$$

$$F|_{(S^*, 0, 0, 0)} = \begin{pmatrix} 0 & \beta S^* \\ 0 & 0 \end{pmatrix}$$

$$V|_{(S^*, 0, 0, 0)} = \begin{pmatrix} \kappa & 0 \\ -\kappa & \gamma \end{pmatrix}$$

$$\text{e) } V^{-1} = \begin{pmatrix} \frac{1}{\kappa} & 0 \\ \frac{1}{\gamma} & \frac{1}{\gamma} \end{pmatrix}$$

$$FV^{-1} = \begin{pmatrix} \frac{\beta S^*}{\gamma} & \frac{\beta S^*}{\gamma} \\ 0 & 0 \end{pmatrix}$$

$$\text{f) } \rho(FV^{-1}) = \mathcal{R}_0 = \frac{\beta S^*}{\gamma}$$

Max Flow Min Cut Method:

1. Infection Flows Through: I, S, E

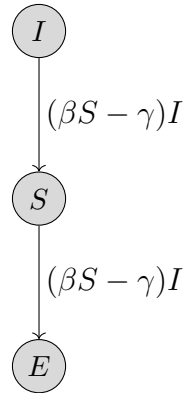
Infection Does Not Flow Through: R

2. $s = I$

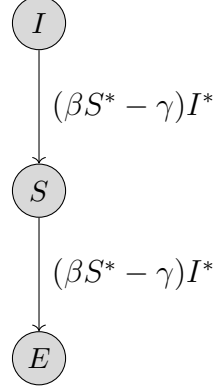
3. $t = E$

4. $\Phi = \{(I, S)\}, w(\Phi) = (\beta S - \gamma)I$

$\Psi = \{(S, E)\}, w(\Psi) = (\beta S - \gamma)I$



5. Evaluate near DFE, $(S^*, I^*, 0)$:



6. Cut set: $\{(I, S), (S, E)\}$

7. Minimum cut weight=Maximum flow: $(\beta S^* - \gamma)I^*$

8. $\text{Max flow} > 0 \implies (\beta S^* - \gamma)I^* > 0 \implies \mathcal{R}_0 = \frac{\beta S^*}{\gamma} > 1$

The set up of the network model in this example is similar to the first and so we still must take into account the net flow of new disease (hence the γ term in the edge weights). The obvious difference is the extra node. Now the sink is the exposed class and the susceptible class is an intermediary node. Following the algorithm, the exposed class is the sink since it feeds directly into the infectious class (source) in the compartmental model. Since the network is set up such that the amount of infection that flows to the sink is the amount available for the source in the next generation and the sink and source cannot be connected, the κ term is ignored in the network.

2.4 Discussion

In addition to calculating the basic reproductive number, using the MFMC method provides valuable information about how we might respond to or manage an outbreak by studying the cut sets that come from the algorithm. For the first example, since the model does not have much complexity, there is only one cut set, $\{(I, S)\}$. Hence the only way for disease to move through the system is directly from the source (infectious) to the sink (susceptibles) and so the only way it can be prevented is by separating the infectious from the susceptibles. Since there is only one cut set, its weight, $(\beta S^* - \gamma)I^*$, is the minimum and

thus is the max flow, which represents the maximum amount of disease to flow through the network that would be available to produce secondary infections in the next generation of infectious individuals. Similarly for the NGM method, our ‘matrix’ is just a number, $\beta S^*/\gamma$, which represents the greatest (maximum) or expected number of secondary infections in the infectious compartment produced by individuals initially in the infectious compartment.

In the second example we have two cut sets: $\{(I, S)\}, \{(S, E)\}$. Even though there is still only one path from the source to the sink ($I \rightarrow S \rightarrow E$), there are two places where the flow could be stopped. The first is to separate the infectious from the rest of the community, $\{(I, S)\}$. In real life this is done by isolating the sick, so they can no longer spread the disease to the susceptible people in the community. The second cut set $\{(S, E)\}$, looks at quarantining anyone that has been exposed to the disease before they become infectious and thus preventing them from causing any new infections. In practice these cut sets recommend similar, yet critically distinct courses of action: quarantine and isolation. In theory, the second cut set represents a better option for disease control because if a person is quarantined before they become infectious they won’t be able to spread the disease, but if a person is isolated only after they are infectious they could have already spread the disease to others before isolation begins. Although better in theory, this option is much harder to implement in practice as contact tracing has many real world limitations. Since both cuts occur on the only path that connects the source and the sink and no disease is lost along the path, they both have the weight of $(\beta S^* - \gamma)I^*$. Thus, this is the max flow and its interpretation is the same as in the first example. Using the NGM method, we note that the expected number of secondary infections in compartment E produced by individuals initially in compartment E and the expected number of secondary infections in compartment E produced by individuals initially in compartment I represent the same idea as the weights of the cut sets and are thus both $\beta S^*/\gamma$. Then since all the other entries of FV^{-1} are 0, $\beta S^*/\gamma$ is the expected number of secondary infections in the next generation.

Beyond the direct method for the simple cases presented, natural extensions of the MFMC method can be used for more complicated models (see Section 2.5 for some example

extensions to more complicated systems; of course, extensions to the basic model to increase capability to capture more complicated dynamics do incur additional complexity in calculation such that they may again become prohibitive without training). This demonstrates the flexibility and power of this method, allowing more natural steps without compromising robustness in cases in which the NGM method is viable yet complicated, but also allowing us to consider some cases in which the assumptions for the NGM method are not met.

2.5 Additional Examples

We can expand the MFMC method to work for models with more complex dynamics. Let $N = S + E + I + R$ and μ be the birth and death rate. Consider the following *SEIR* model with vital dynamics,

$$\begin{aligned}\dot{S} &= \mu N - \mu S - \frac{\beta SI}{N} \\ \dot{E} &= \frac{\beta SI}{N} - (\mu + \kappa)E \\ \dot{I} &= \kappa E - (\gamma + \mu)I \\ \dot{R} &= \gamma I - \mu R.\end{aligned}\tag{2.3}$$

The NGM method tells us that $\mathcal{R}_0 = \frac{\kappa}{\mu + \kappa} \frac{\beta}{\mu + \gamma}$. This basic reproductive number has two components: $\frac{\kappa}{\mu + \kappa}$ and $\frac{\beta}{\mu + \gamma}$. The first factor is the probability of surviving the exposed class and going on to the infectious class and the second factor is the probability of transmitting infection through contact during the infectious period. The MFMC method easily computes the second factor by adjusting the edge weights to include the μ term from the I class, as it functions the same as the removal to the R class, γ , term. The MFMC method must be appended to recover the first factor. Note, since the MFMC method is based on a flow network it does not allow for flow (disease) to leave the sink or, due to the conservation of flow, escape the network. Thus, calculating how much diseases ‘survives’ the sink to be available to the source for the next generation must happen outside of the flow

network. Luckily, determining this survival probability comes straight from the differential equation for the sink class. It has the form:

$$\frac{\text{rate of transition to source}}{\text{rate of transitioning out of sink}},$$

where transition out of the sink includes transition into the source. Thus, multiplying the result of the MFMC method by this fraction recovers $\mathcal{R}_0$. Note, this adjustment is consistent with the previous examples as without death terms the probability of surviving the sink is 1.

Next we will compute the basic reproductive number for (2.3) using the MFMC method with the adjustment described above.

Max Flow Min Cut Method:

1. Infection Flows Through: I, S, E

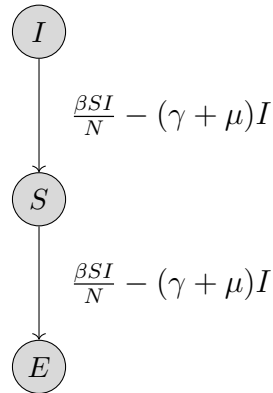
Infection Does Not Flow Through: R

2. $s = I$

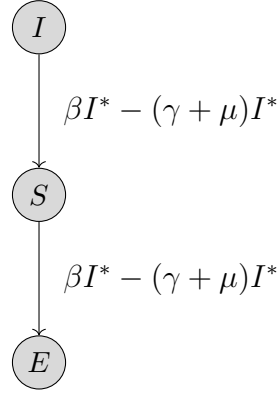
3. $t = E$

4. $\Phi = \{(I, S)\}, w(\Phi) = \frac{\beta SI}{N} - (\gamma + \mu)I$

$\Psi = \{(S, E)\}, w(\Psi) = \frac{\beta SI}{N} - (\gamma + \mu)I$



5. Evaluate near DFE, ($S \approx N, I^* > 0, 0$):



6. Cut set: $\{(I, S), (S, E)\}$

7. Minimum cut weight=Maximum flow: $\beta I^* - (\gamma + \mu) I^*$

8. Max flow $> 0 \implies \beta I^* - (\gamma + \mu) I^* > 0 \implies \frac{\beta}{\gamma + \mu} > 1$

9. Probability of surviving $E = \frac{\text{rate of transition to I}}{\text{rate of transitioning out of E}} = \frac{\kappa}{\mu + \kappa} \implies \mathcal{R}_0 = \frac{\kappa}{\mu + \kappa} \frac{\beta}{\gamma + \mu}$

Example: Rabies in Dogs

Now that we have shown the MFMC method works for a general $SEIR$ model with vital dynamics and a constant population, let's apply the method to an example from the literature. Consider the $SEIR$ model for the spread of rabies in dogs from [76]. It is very similar to the general model with a few additional terms to take into account aspects such as vaccination and disease-related deaths. The model is as follows,

$$\begin{aligned}
 \dot{S} &= A + \lambda R + \sigma(1 - \gamma)E - \beta SI - (m + k)S \\
 \dot{E} &= \beta SI - \sigma(1 - \gamma)E - \sigma\gamma E - (m + k)E \\
 \dot{I} &= \sigma\gamma E - (m + \mu)I \\
 \dot{R} &= k(S + E) - (m + \lambda)R,
 \end{aligned} \tag{2.4}$$

where A is the annual crop of newborn puppies, λ is loss of vaccination immunity, γ is the risk of clinical outcome in exposed dogs, σ is the rate of disease progression, m is the natural death rate, β is the transmission rate, k is the vaccine rate, and μ is the disease specific death rate. The authors of [76] find $\mathcal{R}_0 = \frac{\sigma\gamma}{m+k+\sigma} \frac{\beta S^*}{m+\mu}$, where $S = S^*$ at the DFE. Now we will find $\mathcal{R}_0$ using the adjusted MFMC method defined above.

Max Flow Min Cut

1. Infection Flows Through: I, S, E

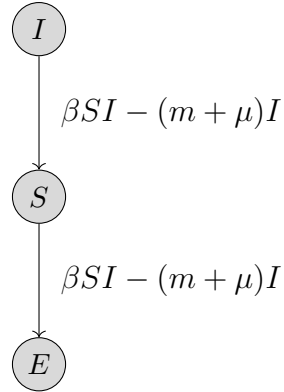
Infection Does Not Flow Through: R

2. $s = I$

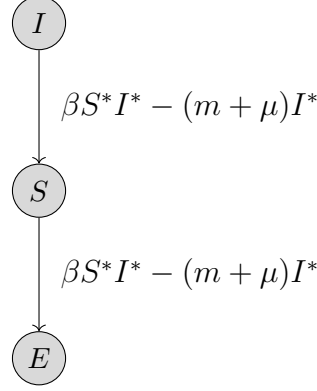
3. $t = E$

4. $\Phi = \{(I, S)\}, w(\Phi) = \beta SI - (m + \mu)I$

$\Psi = \{(S, E)\}, w(\Psi) = \beta SI - (m + \mu)II$



5. Evaluate near DFE, $(S = S^*, I = I^*0, 0)$:



6. Cut set: $\{(I, S)\}, \{(S, E)\}$

7. Minimum cut weight=Maximum flow: $\beta S^* I^* - (m + \mu) I^*$

8. Max flow $> 0 \implies \beta S^* I^* - (m + \mu) I^* > 0 \implies \frac{\beta S^*}{\gamma + \mu} > 1$

9. Probability of surviving $E = \frac{\text{rate of transition to I}}{\text{rate of transitioning out of E}} = \frac{\sigma\gamma}{\sigma(1 - \gamma) + \sigma\gamma + m + k}$
 $= \frac{\sigma\gamma}{\sigma + m + k} \implies \mathcal{R}_0 = \frac{\sigma\gamma}{\sigma + m + k} \frac{\beta S^*}{m + \mu}$

Example: Ebola in Humans

Next let's look at a slightly more complicated model for the literature. Consider the *SEIRD* model for Ebola dynamics from [94]. This model is very similar to a traditional *SEIR*, but includes a class of individuals who have died from Ebola and their bodies are still infectious. The model is as follows,

$$\begin{aligned}
 \frac{dS}{dt} &= -\frac{\beta_I SI}{N} - \frac{\beta_D SD}{N} \\
 \frac{dE}{dt} &= \frac{\beta_I SI}{N} + \frac{\beta_D SD}{N} - \frac{E}{T_E} \\
 \frac{dI}{dt} &= \frac{E}{T_E} - \frac{I}{T_I} \\
 \frac{dR}{dt} &= \frac{(1-f)I}{T_I} \\
 \frac{dD}{dt} &= \frac{fI}{T_I} - \frac{D}{T_D},
 \end{aligned} \tag{2.5}$$

where $N = S + E + I + R$, D is the class of dead, but infectious individuals, β_I is the transmission rate for contact with infected individuals, β_D is the transmission rate for contact with dead, but infectious individuals, T_E is the average exposed period, T_I is the average infectious period, T_D is the average period of infectiousness after death, and f is the fraction of infectious individuals who die and enter the D class. [94] states that $\mathcal{R}_0 = \beta_I T_I + f \beta_D T_D$. Now we will show how to find $\mathcal{R}_0$ using the MFMC method.

Note that this model has two paths to infection: contact with the I class or contact with the D class. These two paths are not in competition and do not have any effect on the other path. Thus, we will formulate two flow systems and compute a basic reproductive number for infections from I , which we will call $\mathcal{R}_I$ and infections from D , which we will call $\mathcal{R}_D$. Then $\mathcal{R}_0$ for (2.5) is $\mathcal{R}_I + \mathcal{R}_D$.

First we will calculate $\mathcal{R}_I$:

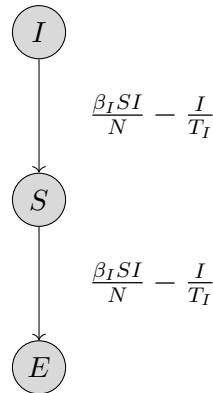
1. Infection Flows Through: I, S, E

Infection Does Not Flow Through: R

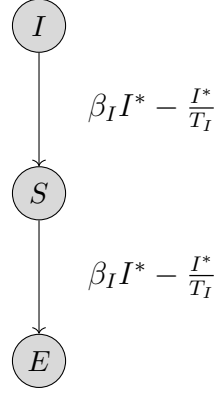
2. $s = I$

3. $t = E$

4. $\Phi = \{(I, S)\}, w(\Phi) = \frac{\beta_I S I}{N} - \frac{I}{T_I}$
 $\Psi = \{(S, E)\}, w(\Psi) = \frac{\beta_I S I}{N} - \frac{I}{T_I}$



5. Evaluate near DFE, ($S \approx N, I^* > 0, 0$):



6. Cut set: $\{(I, S), (S, E)\}$

7. Minimum cut weight=Maximum flow: $\beta_I I^* - \frac{I^*}{T_I}$

8. Max flow $> 0 \implies \beta_I I^* - \frac{I^*}{T_I} > 0 \implies \mathcal{R}_I = \beta_I T_I > 1$

Now we will calculate $\mathcal{R}_D$:

1. Infection Flows Through: D, S, E

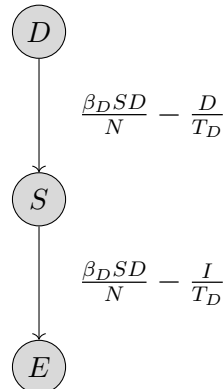
Infection Does Not Flow Through: $\emptyset$

2. $s = D$

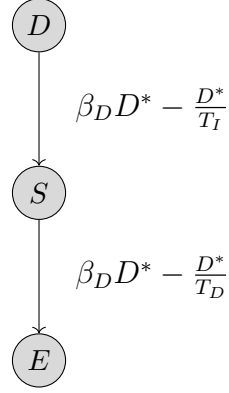
3. $t = E$

4. $\Phi = \{(D, S)\}, w(\Phi) = \frac{\beta_D S D}{N} - \frac{D}{T_I}$

$\Psi = \{(S, E)\}, w(\Psi) = \frac{\beta_I S D}{N} - \frac{D}{T_D}$



5. Evaluate near DFE, $(S \approx N, D^* > 0, 0)$:



6. Cut set: $\{(D, S)\}, \{(S, E)\}$

7. Minimum cut weight=Maximum flow: $\beta_D D^* - \frac{D^*}{T_D}$

8. Max flow $> 0 \implies \beta_D D^* - \frac{D^*}{T_D} > 0 \implies \beta_D T_D > 1$

9. Probability of transitioning to D : $f \implies \mathcal{R}_D = f\beta_D T_D$

Therefore, $\mathcal{R}_0 = \mathcal{R}_I + \mathcal{R}_D = \beta_I T_I + f\beta_D T_D$. Note that step 9. is applied when calculating $\mathcal{R}_D$. Just as in (2.3) where we have to calculate, outside the network, how much disease ‘survives’ E to go on to I in the next generation, in (2.5) we have to calculate how much disease ‘survives’ the infectious class (transitions to D) to be the source of infection for the $D - S$ infection route. Further this step is needed since the population is not constant, the definition of N does not include D , so we have to consider the amount of ‘disease’ that leaves the network. We do this by considering the probability an infectious person dies and transitions out of the population into D . By taking that probability into account, we capture the amount of infection that comes from disease that left the network.

2.6 Conclusion

While the NGM method is a robust and widely used method for computing the basic reproductive number, the MFMC method is a useful alternative. By relying solely on a

networks approach, this opens a new perspective for thinking and learning about disease and transmission dynamics. Especially as epidemiological results often must be presented to, and understood by, a non-mathematical audience, this graphical approach is easier to explain and requires less foundational mathematical knowledge to understand. However, in addition to providing a more accessible, alternative approach to understanding the basic reproductive number, the MFMC method also gives insights that the NGM method does not. As discussed above, the MFMC method offers information about the dynamics of the system and different methods of preventing an outbreak from progressing by analyzing the cut sets. More intriguingly, however, it can also be applied to systems that have non-unique endemic equilibria and even provide non-equilibrium behavior estimates by employing continuous time dynamics [67] and continuous dynamics with delays [66] in the flow network. By alleviating the constraints imposed by having to solve by substitution at an equilibrium point, these methods can allow much more nuanced characterization of transient behavior.

Considering a well-studied problem from a different angle can allow for new insights. We have developed a network theoretic approach for calculating and analyzing the basic reproductive number and demonstrated how the MFMC method is an equivalent method to the traditional NGM method. By analyzing the cut sets of infection flow, we show how the MFMC method can provide valuable, conceptually intuitive information about how to stop an outbreak from progressing. Beyond simplifying access to stable system dynamics, extensions of the methods presented also provide additional tools for analyzing transient behaviors. The presented novel MFMC method increases our epidemiological toolkit for teaching and explaining disease dynamics and the basic reproductive number. The availability of this more accessible and intuitive methodology can expand our ability to communicate the perspectives and insights of mathematical modeling to policy makers and help enable real-time response to incipient outbreaks.

Chapter 3

Linking Immuno-Epidemiology Principles to Violence

Publication information: In review

Disclosure statement: Anna Sisk is the first and corresponding author on the manuscript based on this chapter. Her co-authors are Nina Fefferman (advisor), Judy Day (committee member), and Patricia Bamwine. Fefferman and Day conceived and directed the work. Sisk implemented the analysis, wrote the code, and drafted the manuscript. All the authors contributed to the formulation of the model, analogies/framework, and model analysis. All the authors worked to review, revise, write, and approve the manuscript.

Abstract

Societies have always struggled with violence, but recently there has been a push to understand violence as a public health issue. This idea has unified professionals in medicine, epidemiology, and psychology with a goal to end violence and heal those exposed to it. Recently, analogies have been made between community-level infectious disease epidemiology and how violence spreads within a community. Experts in public health and medicine suggest an epidemiological framework could be used to study violence.

Building upon results from community organizations, which implement public health-like techniques to stop violence spread, we look to formalize the analogies between violence and infectious diseases. Then expanding on these ideas and using mathematical epidemiological principals, we formulate a susceptible-exposed-infected model that captures violence spread. Further, we ran example numerical simulations and performed equilibrium and stability analysis to show how a mathematical model can provide insight on prevention strategies.

The simulations show negative effects of violence exposure have a greater impact than positive effects of preventative measures. By halving the impact of violence exposure, the amount of violence in a community drastically decreases in the long-term; but to reach this same outcome through an increase in the amount of after exposure support, it must be approximately 5-fold. We note that our simulations qualitatively agree with empirical

studies. Further, we build upon the work in Chapter 2 to expand upon the basic reproductive number in a way that is useful for our model. Having a mathematical model gives insights on the effectiveness of different strategies for violence prevention. Our simulations suggest the most effective use of community funding is investing in protective factors, instead of support after violence exposure.

3.1 Introduction

Societies have always struggled with the causes and effects of violence and to a large extent, current methods for curbing violence and its detrimental ripple effects in communities are inadequate. Unfortunately, children and adolescents are often disproportionately impacted by these inadequacies [64]. Each day in the United States, more than 10 adolescents are victims of homicide and over 1,300 are treated for non-fatal assault-related injuries, according to the Centers for Disease Control and Prevention (CDC) [22]. This age group is at a greater risk for poly-victimization, which is experiencing multiple forms of victimization and/or violence. According to The National Survey of Children’s Exposure to Violence, a child who has experienced a physical assault would be five time more likely to have been sexually abused and more than four times as likely to have experienced maltreatment [34]. Furthermore, while the effects of violence reach every demographic, young people in minority communities bear the brunt of the consequences. Homicide is the leading cause of death for Black adolescents and young adults aged 15 to 24 and the second leading cause of death for Hispanic adolescents and young adults in the same age range [93]. The primary way to handle the epidemic of violence our society faces is through the criminal justice system, but recently there has been a drive to treat violence as a disease. With this new view of violence, we can utilize community-based intervention models to better understand and treat violence from a public health perspective.

One recently developed approach focuses on viewing violence as an infectious disease. This idea was based on similarity to patterns of spread in maps of disease and violence outbreaks, such as the appearance of clusters that easily spread to surrounding areas [81].

The intervention-focused organization, CeaseFire, in Chicago, Illinois attempts to prevent violence in high crime areas by implementing intervention/disruption techniques similar to those used to prevent disease spread, such as using contact tracing to find people who are exposed to an act of violence, developing community awareness programs, and both physically and socially distancing individuals at risk of engaging in violence. Thus far, these efforts have achieved a decrease in all shots fired in 4 of the 7 neighborhoods in which CeaseFire has operated [80]. Community-based violence intervention programs can now be seen in American cities like New York, Baltimore, Pittsburgh, Kansas City, Philadelphia, and Oakland, along with international programs in Colombia, the United Kingdom, Mexico, Canada, and South Africa [43].

These programs provide preliminary empirical evidence that treating violence as an infectious disease may be an effective approach for violence prevention. This suggests that further efforts to employ the tool kits of infectious disease prevention may also be beneficial. In epidemiology, mathematical modeling of infectious disease dynamics helps complement our understanding from experimental studies. For instance, the susceptible-infected-recovered (SIR) epidemiological model and its variants, have been mathematically modeled to predict how a disease will spread and to test the effectiveness of different interventions, such as vaccines. A recent example of where this work is clearly showcased is the COVID-19 pandemic [6, 42, 79]. Thus, experts in public health and medicine have suggested that these approaches could be used to study violence [81].

Violence and crime, while connected, are distinct phenomena. Note, that not all crime is violent (for example: fraud, drug possession, petty theft, etc.), and not all violence is legally criminal/reported as a crime (for example: bullying, suicide, intimate partner violence, etc.). Crime, by its definition, is a violation of the law, while violence can be defined as physical methods by which to harm others, even if the harm they produce is non-physical (i.e. social harm, deprivation of resources, etc.) [33]. Thus, even though crime and violence are clearly connected, they can have different causes [33], so it is important that each of these phenomena are studied separately. While mathematical and statistical models have been employed to

study the spread of crime, riots, and policing in the past [58, 12, 10, 84, 73, 86, 96], more investigation is needed in exploring exposure-driven violence, detached from criminality and policing. Recently, agent-based modeling has been used to show the effectiveness of community-based violence prevention organizations, like Cure Violence (formerly known as CeaseFire) [16]. Although, little work has been done to model violence spread using mathematical epidemiological principles.

One yet unexplored, class of mathematical models that may be useful for studying violence is Immuno-Epidemiological models. These are mathematically dynamic multi-scale models that link a within-host disease model to an epidemiological population-level model [60]. This integrative approach allows for the bidirectional feedback between the individual and population level to be studied. It is the ideas from this model type we want to implement to study violence. Specifically, we use epidemiological models to explore the spread of violence on the community level and an immune system model to look at the impact that exposure to violence has at the individual level (as described by the change in their propensity to later commit violent acts themselves). By looking at violence through the Immuno-Epidemiological modeling lens we can gain a deeper understanding of the connection between exposure to violence, prevention and mitigation techniques, and the threshold for committing a violent act. Further, by using this model we will be able to quickly and efficiently explore different violence prevention methods without needing to implement costly and time-consuming empirical studies.

To achieve these goals, we first expand on the initial analogy of violence as an infectious disease by incorporating the individual (host) scale into the framework and formalize the analogy at both scales. We do not suggest this as a pathway by which to address all of the many varied potential causes of, or methods by which to address, violence (e.g., patriarchy, institutional inequalities, intergenerational trauma, etc.), but this does not mean these methods are without targeted and practical use. Even though mathematical epidemic models do not focus on “solving” the root cause of infectious outbreaks (germs), they still provide critical tools by which to plan mitigative strategies and effective interventions. In

this same way, we are proposing a quantitative way to analyze the dynamics of violence spread that occurs through exposure among young people, and the relative effectiveness of potential intervention strategies. In this chapter we focus on building the mathematical structure for an epidemiological model that describes violence spread through exposure. We use this epidemiological model for violence to run some sample simulations and compare the qualitative outcomes of these simulations to the results from in-field studies. Further, we conduct equilibrium and stability analysis and look at the role of a basic reproductive number in a epidemiologically based violence model. These results help us, not only confirm the mathematical epidemiological approach to studying violence is realistic and useful, but also explore some general dynamics between violence, exposure, and intervention/disruption techniques and protective factors.

3.2 Methods

3.2.1 Analogy

To continue building on the idea of violence as an infectious disease, we formalize and expand on this analogy. One of the main ways we did this is to include a within-host component. In the fields of psychology, sociology, and social work, it is well known that exposure to violence not only increases an individual’s risk for further victimization, but also for committing violence themselves, and that the use of intervention/disruption techniques (such as those implemented by CeaseFire) and protective factors (which are resources that help mitigate the effects of risk factors/violence exposure [64]) can reduce propensity to commit violence [39, 47]. Based on these established and understood principles, we build our model to rely explicitly and uniquely on these factors. Thus, we consider that exposure to violence can have deleterious effects on a person’s mental health just as exposure to an agent of disease can lead to an assault on one’s physical health. Similarly, as intervention/disruption techniques and protective factors can positively influence mental health, a person’s immune system can provide intervention/disruption to the disease progression and promote host health. For example, protection and promotion of host health can be achieved through

intervention/disruption techniques for disease spread mitigation such as social distancing, contact tracing, vaccines, medical treatment and prevention, increased sanitation, and awareness programs. Then following the analogy, protection and promotion of mental health can be achieved through interruption/disruption techniques like those mentioned previously or protective factors, such as family and school connectedness, economic relief in forms of tax credits, access to affordable childcare, nutritional assistance programs, mentoring programs for both youths and parents, and therapeutic support [39, 36].

As with any model of a real-world system, some simplifying assumptions are made to build a mathematical representation for the analogy between violence and infectious disease dynamics. For instance, while there are many different forms of violence (verbal/emotional/physical/etc.) we are focusing on physical, reportable acts of violence. This includes, but is not limited to, physical or sexual assaults/abuse, suicide or suicide attempts, homicide or homicide attempts, gun violence, physical bullying, robberies/muggings, or intimate partner violence. Thus, this analogy is not restricted to looking at one specific type of violence; instead, we are interested in exploring the role that poly-victimization plays in violence spread. This model is designed with the assumption it will be representing youths aged approximately 14-17 without severe underlying mental health issues where the illness can manifest as acts of violence, as we are focusing on the role that exposure plays in violence propagation. Below the model is described along with the associated assumptions used to build it.

3.2.2 Model Description

The susceptible-exposed-infectious (SEI) model is a well-regarded epidemiological construct that divides a population into classes: a susceptible class, exposed class, and infectious class. In traditional disease modeling the susceptible class is defined to be the group of individuals who are not infected with the disease, but are susceptible to becoming infected and the infectious class is the individuals who are infected with the disease and importantly, are capable of transmitting it to others in the population. For the violence spread model, the susceptible class is a naive group of individuals who have never been exposed to violence;

the exposed class is the group who have had exposure to violence, but are not committing violence themselves; and the infectious class are those who are actively committing violence. The most basic way for movement between the classes to occur is from interactions between the classes or interaction with an act of violence. Mathematically this is expressed by modeling all interactions through typical mass action terms, which makes the simplifying assumption that all members of each class are able to interact at any time. Two versions of the SEI model for violence have been created: one that considers only a single exposed class (One Exposed Class Model), (3.1), and another that includes three different exposed classes (Three Exposed Class Model), (3.2).

One Exposed Class Model:

$$\begin{aligned}
 \dot{S} &= -\mu S(t)I(t) \\
 \dot{E} &= \mu S(t)I(t) - \omega E(t)I(t) + \beta I(t) \\
 \dot{I} &= \omega E(t)I(t) - \beta I(t)
 \end{aligned} \tag{3.1}$$

Unlike the One Exposed Class Model, the Three Exposed Class Model divides the general exposed population into three exposure sub-classes: Exposed₁, Exposed₂, and Exposed₃. The exposed classes are defined in the following way: those in the Exposed₁ class have violence exposure that causes the smallest increase in propensity to commit violence, those in the Exposed₂ have violence exposure that causes an intermediate increase in propensity to commit violence, and those in the Exposed₃ class have violence exposure that causes the most severe increase in propensity to commit violence. Further, the third exposed class is at the highest risk to move into the infectious class, which is committing an act of violence. In the One Exposed Class Model, any exposure to violence would lead straightaway to being in the highest (and only) risk class for moving into the infectious class.

Three Exposed Class Model:

$$\begin{aligned}\dot{S} &= -\mu_1 S(t)I(t) - \mu_2 S(t)I(t) - \mu_3 S(t)I(t) \\ \dot{E}_1 &= \mu_1 S(t)I(t) - \omega_1 E_1(t)I(t) - \omega_2 E_1(t)I(t) + \beta_1 E_2(t) \\ \dot{E}_2 &= \mu_2 S(t)I(t) - \omega_3 E_2(t)I(t) - \beta_1 E_2(t) + \omega_1 E_1(t)I(t) + \beta_2 E_3(t) \\ \dot{E}_3 &= \mu_3 S(t)I(t) - \omega_4 E_3(t)I(t) - \beta_2 E_3(t) + \omega_2 E_1(t)I(t) + \omega_3 E_2(t)I(t) + \beta_3 I(t) \\ \dot{I} &= \omega_4 E_3(t)I(t) - \beta_3 I(t)\end{aligned}\tag{3.2}$$

Further, it is assumed that the population is closed, meaning the size of the population is unchanging over time. For example, births and deaths in the system are considered to be balanced and therefore ignored. This means the total population, denoted N , is equal to the sum of all the classes. Mathematically, this means $N = S + E + I$ for the One Exposed Class Model and $N = S + E_1 + E_2 + E_3 + I$ for the Three Exposed Class Model, where S is the susceptible class, I is the infectious class, E is the exposed class, E_1 is the first exposed class, E_2 is the second exposed class, and E_3 is the third exposed class. For the purposes of this text the units of time are arbitrary, but in the future the parameters will be calibrated to have time units of months based on some preliminary review of available data. For different assumptions on time the dynamics can be calibrated accordingly with the available data. A further assumption made is about contact between the classes. In traditional disease modeling, contact usually implies physical or face to face interactions between the classes; however, the proposed model does not require physical interaction between the classes and “exposure to” or “contact with” violence could be direct or indirect. For example, a person who is exposed to violence could mean they are the victim (direct), or the victim is in the person’s close personal network (indirect), or even a violent act occurred in a person’s community (indirect). The final important assumption made is those in the infectious class are actively committing violence. Thus, once the act of violence has ended, there is movement out of the infectious class back to the exposed class(es).

3.2.3 Model Parameters

For the One Exposed Class Model the parameter μ governs the effect of initial exposure to violence, representing the rate that susceptible individuals become exposed to violence. Next, ω is the rate at which violence occurs. This parameter could be estimated by looking at the effectiveness of the intervention/disruption techniques. If these are working well then ω would be small; if not, it would be large. It could also be estimated by looking at the literature to determine how poly or re-victimization affects a person's mental health and, thus, propensity to commit violence. Since this parameter is linked to an individual's behavior and mental health, this is where a within-host level model, discussed earlier, could come into play. Finally, β acts as an automatic removal rate, since once the actual act of violence ends there is immediate removal back to the exposed class. Figure 3.1 shows the diagram of the One Exposed Class Model with the arrows showing movement between the classes and the parameters representing the rates that the movements occur.

While the parameters in the Three Exposed Class Model are similar to those in the One Exposed Class Model, they have important differences. Note, β_1 and β_2 are related to the effectiveness of intervention/disruption techniques, protective factors, and the healing that comes with the passage of time. The more effective these techniques are the faster an individuals' propensity to commit violence will decrease and the faster people will move to a/the less affected exposed classes. While β_3 does move people back one class like the other β parameters, it does not rely on the effectiveness of the intervention/disruption techniques or protective factors like the others do. Once an act of violence is finished, the aggressor immediately moves back to Exposed₃ independent of everything. Thus, this parameter functions more as an automatic removal rate than a recovery rate that is dependent on interventions.

The ω_i parameters look at the effects of exposure to violence, rather than just the exposure itself. Since this is the parameter that moves people from one exposed class to the next, it needs to take into account how exposure to violence affects a person's propensity to commit

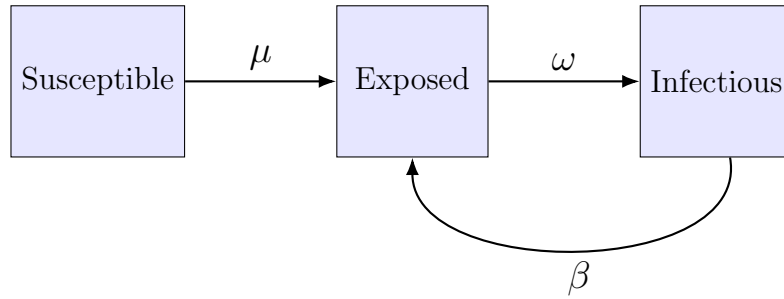


Figure 3.1: Compartment Diagram for One Exposed Class Model. This is formulated from an SEI epidemiological model to describe violence spread, where the susceptible class has never been exposed to violence; the exposed class has had exposure to violence, but are not committing violence; and the infectious class are actively committing violence.

violence. Thus, not every exposure to violence will move an individual to the next exposed class, only exposures that have a significant enough impact to increase the likelihood of committing violence. These parameters should include two components: one that looks at the ‘quality’ of the exposure (direct or indirect) and the other that takes into account the robustness of a person’s mental health to an exposure. Thus, with the right support system in place exposure to a certain act of violence would have less of an effect on a person’s likelihood to commit violence than if that support is absent. Just as in the One Exposed Class Model, since the ω_i parameters are focused on an individual’s response to violence exposure, this is a possible place for a within-host model to come into play. The functional form for this parameter could be described as an exponential function of number of exposures, since the literature reveals that after initial exposure to violence, the likelihood of re-victimization or poly-victimization increases [64]; and increased exposure to violence increases the probability of later committing violence [47, 39]. Thus, the values of the ω_i ’s are taken with $\omega_1 < \omega_3 < \omega_4$ according to an exponential relationship rather than a linear one, for instance. Finally, μ_1, μ_2 , and μ_3 focus on the impact that an initial exposure to violence has on a person’s tendency to later commit violence, and essentially provides a way for susceptibles to ‘skip passed’ one (μ_2) or two (μ_3) exposed classes if, for instance, the initial exposure is particularly influential. Table 3.1 compares the interpretation of the parameters for the Three Exposed Class Model to their traditional epidemiological interpretations and Figure 3.2 shows the diagram of the model.

3.2.4 Equilibrium and Stability Analysis

We start exploring the dynamics of both models by conducting equilibrium and stability analysis. To do this we found the equilibrium points of the system, which are solutions to the system that are constant. We set each equation in the system equal to zero and solved for the state variables, S, E, I for the One Exposed Class Model or S, E_1, E_2, E_3, I for the Three Exposed Class Model. Equilibrium points are given as ordered tuples. The notation we use is (Susceptible, Exposed, Infected)/(Susceptible, Exposed₁, Exposed₂, Exposed₃, Infected) respectively. In traditional disease modeling there are two type of equilibrium points: disease free and endemic. For our adaptation to violence modeling

Table 3.1: Parameters. Lists the parameters for the Three Exposed Class Model, where the second column is the traditional epidemiological definitions and the third column describes the parameters as applied to violence.

Name	Epidemiological	Violence
μ_1	initial exposure rate that results in a low likelihood of transmission	rate at which initial exposure to violence increases likelihood of later committing violence (low increase)
μ_2	initial exposure rate that results in a intermediate likelihood of transmission	rate at which initial exposure to violence increases likelihood of later committing violence (intermediate increase)
μ_3	initial exposure rate that results in a high likelihood of transmission	rate at which initial exposure to violence increases likelihood of later committing violence (high increase)
ω_1	re-exposure rate	rate at which exposure to an act of violence increases the probability of committing violence
ω_2	re-exposure rate	rate at which exposure to an act of violence increases the probability of committing violence (greater increase then ω_1)
ω_3	re-exposure rate	rate at which exposure to an act of violence increases the probability of committing violence
ω_4	transmission rate	rate at which an exposure to violence leads to committing an act of violence
β_1	recovery rate	decrease in the propensity to commit violence
β_2	recovery rate	decrease in the propensity to commit violence
β_3	removal rate	cessation (of violence) rate

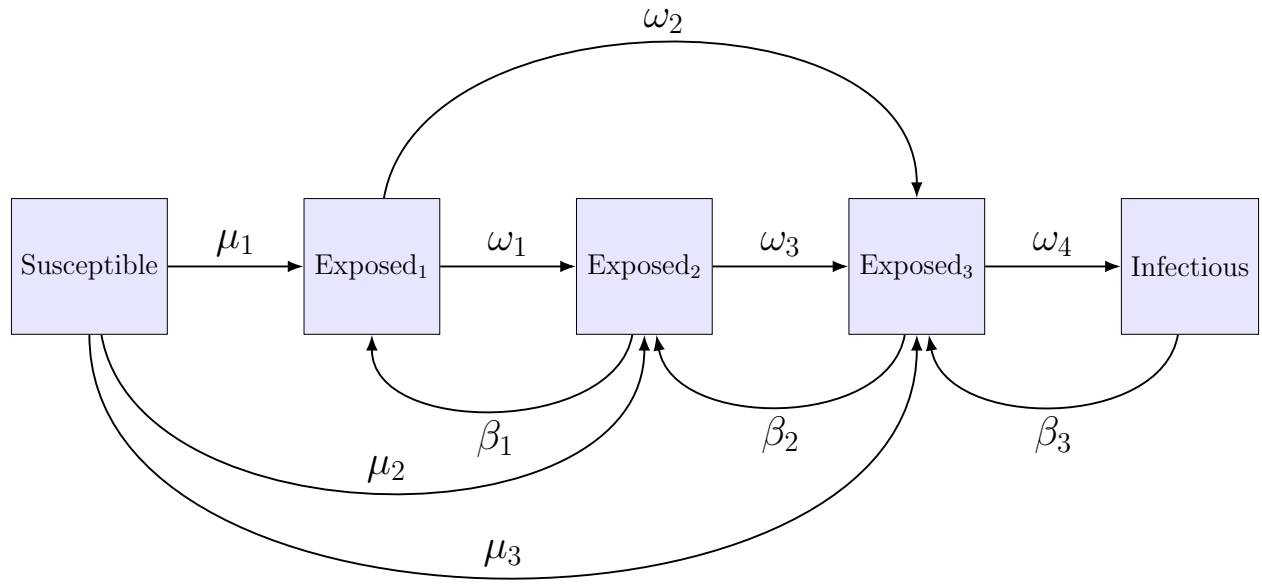


Figure 3.2: Compartment Diagram for Three Exposed Class Model. It is very similar to Figure 3.1, with the exception of having three exposed classes and the additional associated parameters. The Exposed_1 class has had violence exposure that causes the smallest increase in propensity to commit violence, the Exposed_2 class has had violence exposure that causes an intermediate increase in propensity to commit violence, the Exposed_3 class has had violence exposure that causes the most severe increase in propensity to commit violence.

we have three types of equilibrium points: violence free equilibrium (VFE), semi-violence free equilibrium (SVFE), and endemic equilibrium. The method we used for determining stability is linearization, which is finding the Jacobian of the system, evaluating the resulting matrix at the desired equilibrium point, then analyzing the eigenvalues of the matrix. If all the eigenvalues are negative the equilibrium point is stable, if they are all positive it is unstable, and if at least one is zero then stability cannot be determined. If they are stable, in the long term the system will go towards that solution and if they are unstable, in the long term the system will repel away from it.

Due to both models having non-unique equilibrium points linearization did not immediately work to determine stability. For the One Exposed Class Model, since the system is closed, we substituted out the S variable using the relationship $N = S + E + I$. This reduced the dimension to two, E and I , so we can use graphical techniques to learn about stability. We first determined the nullclines, which are places in the phase plane where the slopes are zero, by solving $\dot{I} = 0$ and $\dot{E} = 0$ for the state variable. The solutions to the former equation are the I-nullclines and the latter are the E-nullclines. Using this information we then used traditional linearization techniques to determine the stability of the endemic equilibrium point. Then to confirm the results of the stability analysis we use the software Pplane to create a phase portrait. The phase portrait plots trajectories of the system using different initial conditions which gives us an idea of the system's behavior in the long run. The initial conditions and parameters used for the figures were selected at random with the requirement they satisfied any biological feasibility conditions.

Due to the increased complexity of the Three Exposed Class Model, Descartes' Rule of Signs was used to show existence of the endemic equilibrium. Recall that Descartes' Rule of Signs states the number of positive roots of a polynomial is equal to the number of sign changes of consecutive, non-zero coefficients or less than it by an even number. Further, the number of negative roots is equal to the number of sign changes or less than it by an even number when a negative number is substituted into the polynomial. Then traditional linearization

techniques were implemented to attempt to determine stability. For both models, Maple is used to perform the calculations.

3.2.5 Numerical Simulations

To further understand the specific and more complicated dynamics described by the Three Exposed Class Model we simulated some basic scenarios. The output of these simulations confirms the model is displaying expected behavior with respect to the scenarios explored. Thus, we moved forward with exploring additional system behaviors. Using these simulations, we can investigate different example scenarios and gain insight into the feedback between violence exposure, interruption/disruption techniques, and protective factors. To simulate the model scenarios, it was necessary to solve the system of equations, which was done with Python, utilizing the `dopri5` solver. The solutions give the time courses of each class in the model, providing a short- and long-term look at the behaviors of the sub-populations under varying parameter values. For the simulations a population size of 100 individuals was assumed with the initial population distributed into the classes as follows: 10 in the Susceptible class, 50 in Exposed_1 , 30 in Exposed_2 , 5 in Exposed_3 , and 5 in the Infectious class. As described in the previous section, the ω_i parameters were chosen based on an exponential relationship. In addition, the μ_i parameters were made larger than the ω_i parameters, based on an assumption that initial exposure to violence will occur at a faster rate than an increase in the rate of propensity to commit violence after initial exposure. Further, β_3 was chosen to be the largest parameter as it acts as an automatic removal term from the infectious class. These parameter values, though determined heuristically at this stage, enable the exploration of simulations that provide an alternative way to study the effectiveness of various interventions and the outcomes of different parameter combinations and fluctuations over time. Therefore, these proof-of-concept simulations and example scenarios provide a means to see the value in exploring and experimenting with model simulations to ultimately help reduce the overall effect of violence in a community.

3.3 Results

3.3.1 Equilibrium and Stability Analysis

After performing equilibrium analysis on both models we found the VFE, SVFE, and the endemic equilibrium as seen in Table 3.2. Note that the SVFE is the general form of the violence/semi-violence free equilibrium. Thus, there is an infinite amount of semi-violence free equilibrium. Further we note that $(0, 0, 0)/(0, 0, 0, 0, 0)$ are also equilibrium points for the One Exposed Class and Three Exposed Class models respectively, but this would imply that our population is zero, so we ignore it.

One Exposed Class Model

Starting with the One Exposed Class Model, for biological feasible of the endemic equilibrium we require $N\omega - \beta > 0$. If this condition does not hold then the number of people in the infectious class would be negative. To determine the stability of the One Exposed Class Model's equilibrium points we first reduced the three dimensional system to a two dimensional system. Note that $S + E + I = N \implies S = N - E - I$. Making this substitution into (3.1) we get the reduced 2-D system:

$$\begin{aligned}\dot{E} &= \mu(N - E(t) - I(t))I(t) - \omega E(t)I(t) + \beta I(t) \\ \dot{I} &= \omega E(t)I(t) - \beta I(t).\end{aligned}\tag{3.3}$$

Now that we have reduced our system to two dimensions we can find the nullclines, which are listed in Table 3.3 and shown graphically in Figure 3.3. Since $I = 0$ is both an E-nullcline and an I-nullcline we know there are an infinite amount of VFE/SVFE and thus we consider these to be a marginal case of non-isolated equilibrium points. Therefore, the stability of these equilibrium are sensitive to small non-linear terms in the system and cannot be definitively determined. Looking at Figure 3.3 we note the endemic equilibrium is an isolated point and linearization techniques can be used to determine its stability.

In order to determine the stability of the endemic equilibrium we must first calculate the

Table 3.2: Equilibrium Points. $E_1^* = \frac{\beta_3 \beta_1 \beta_2}{I^* \omega_4 (I^* \omega_1 \omega_3 + I^* \omega_2 \omega_3 + \beta_1 \omega_2)}, E_2^* = \frac{\beta_3 \beta_2 (\omega_1 + \omega_2)}{\omega_4 (I^* \omega_1 \omega_3 + I^* \omega_2 \omega_3 + \beta_1 \omega_2)}, I^* > 0$ but small, and $0 \leq S^* \leq N$.

Category	One Exposed Class Model	Three Exposed Class Model
VFE	$(N, 0, 0)$	$(N, 0, 0, 0, 0)$
SVFE	$(S^*, N - S^*, 0)$	$(S^*, N - S^*, 0, 0, 0)$
Endemic Equilibrium	$(0, \frac{\beta}{\omega}, N - \frac{\beta}{\omega})$	$(0, E_1^*, E_2^*, \frac{\beta_3}{\omega_4}, N - E_1^* - E_2^* - \frac{\beta_3}{\omega_4})$

Table 3.3: Nullclines. I and E nullclines for 3.3.

I-Nullcline	E-Nullcline
$I = 0$	$I = 0$
$E = \frac{\beta}{\omega}$	$E + I \frac{\mu}{\omega + \mu} = \frac{N\mu + \beta}{\omega + \mu}$

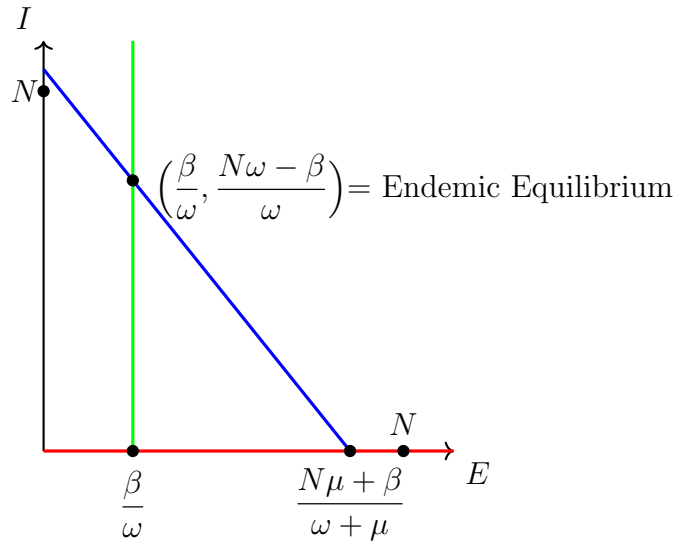


Figure 3.3: Nullcline Plot. Nullcline plot for 3.3, where the red line is the $I = 0$ nullcline, the blue line is the $E + I \frac{\mu}{\omega + \mu} = \frac{N\mu + \beta}{\omega + \mu}$ nullcline, and the green line is the $E = \frac{\beta}{\omega}$ nullcline. The only unique equilibrium point, the endemic equilibrium, is labeled. The red line represents the infinite amount of VFE/SVFE equilibrium.

Jacobian of (3.3), which is,

$$J = \begin{bmatrix} -\mu I - \omega I & \mu N - \mu E - 2\mu I - \omega E + \beta \\ \omega I & \omega E - \beta \end{bmatrix}.$$

Then evaluating J at the endemic equilibrium and simplifying we get,

$$\begin{aligned} J|_{\left(\frac{\beta}{\omega}, \frac{N\omega - \beta}{\omega}\right)} &= \begin{bmatrix} -\mu\left(\frac{N\omega - \beta}{\omega}\right) - \omega\left(\frac{N\omega - \beta}{\omega}\right) & \mu N - \mu\left(\frac{\beta}{\omega}\right) - 2\mu\left(\frac{N\omega - \beta}{\omega}\right) - \omega\left(\frac{\beta}{\omega}\right) + \beta \\ \omega\left(\frac{N\omega - \beta}{\omega}\right) & \omega\left(\frac{\beta}{\omega}\right) - \beta \end{bmatrix} \\ &= \begin{bmatrix} \frac{\mu(\beta - N\omega)}{\omega} + \beta - N\omega & \mu N - \frac{\mu\beta}{\omega} - 2\mu\left(\frac{N\omega - \beta}{\omega}\right) \\ N\omega - \beta & 0 \end{bmatrix}. \end{aligned}$$

Using Maple to find the characteristic equation and eigenvalues we have,

$$\begin{aligned} \det(J|_{\left(\frac{\beta}{\omega}, \frac{N\omega - \beta}{\omega}\right)} - \lambda Id) &= \frac{((\mu N + \lambda)\omega - \beta\mu)(N\omega - \beta + \lambda)}{\omega} = 0 \\ \implies \lambda_1 &= -N\omega + \beta, \lambda_2 = -\frac{\mu(N\omega - \beta)}{\omega}. \end{aligned}$$

Then by the biologically feasible condition we know $\lambda_1 < 0, \lambda_2 < 0$. Thus, the endemic equilibrium is stable. The stability of the endemic equilibrium is further confirmed by its phase portrait, Figure 3.4.

Three Exposed Class Model

Due to the complexity of the endemic equilibria for the Three Exposed Class Model from Table 3.2 and the dependence on I^* it is not obvious what condition(s) need to be satisfied for the existence of a biologically feasible endemic equilibrium. Since we are assuming $I^* > 0$, then clearly $E_1^*, E_2^* > 0$. Further, E_3^* is positive since all the parameters are positive. Now

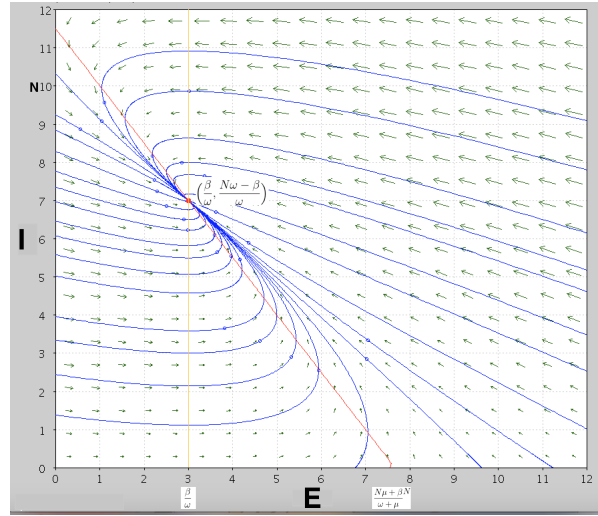


Figure 3.4: Phase Portrait. Phase Portrait for 3.3 drawn with Pplane where $N = 10, \beta = 3, \mu = 2, \omega = 1$. These parameter were picked randomly with the only requirement being they satisfy condition biologically feasible condition.

we want determine what conditions need to be met for I^* to be positive. Observe,

$$\begin{aligned}
I^* &= N - E_1^* - E_2^* - E_3^* \\
I^* &= N - \frac{\beta_3 \beta_1 \beta_2}{I^* \omega_4 (I^* \omega_1 \omega_3 + I^* \omega_2 \omega_3 + \beta_1 \omega_2)} - \frac{\beta_3 \beta_2 (\omega_1 + \omega_2)}{\omega_4 (I^* \omega_1 \omega_3 + I^* \omega_2 \omega_3 + \beta_1 \omega_2)} - \frac{\beta_3}{\omega_4} \quad (3.4) \\
0 &= \omega_4 (\omega_1 \omega_3 + \omega_2 \omega_3) I^{*3} + (-N \omega_1 \omega_3 \omega_4 - N \omega_2 \omega_3 \omega_4 + \omega_2 \omega_4 \beta_1 + \beta_3 \omega_1 \omega_3 \\
&\quad + \beta_3 \omega_2 \omega_3) I^{*2} + (-N \beta_1 \omega_2 \omega_4 + \beta_3 \beta_1 \omega_2 + \beta_2 \beta_3 \omega_1 + \beta_2 \beta_3 \omega_2) I^* + \beta_3 \beta_1 \beta_2
\end{aligned}$$

Let,

$$\begin{aligned}
a &:= \omega_4 (\omega_1 \omega_3 + \omega_2 \omega_3) \\
b &:= -N \omega_1 \omega_3 \omega_4 - N \omega_2 \omega_3 \omega_4 + \omega_2 \omega_4 \beta_1 + \beta_3 \omega_1 \omega_3 + \beta_3 \omega_2 \omega_3 \\
c &:= -N \beta_1 \omega_2 \omega_4 + \beta_3 \beta_1 \omega_2 + \beta_2 \beta_3 \omega_1 + \beta_2 \beta_3 \omega_2 \\
d &:= \beta_3 \beta_1 \beta_2.
\end{aligned}$$

Then (3.4) can be written as,

$$aI^{*3} + bI^{*2} + cI^* + d.$$

Clearly, $a > 0, d > 0$. Let's first consider if $b, c \neq 0$. If we assume that $b, c > 0$, then there are no sign changes, $(++++)$, which implies there are no positive roots by Descartes' Rule of Signs. This would not be biologically feasible. Thus, we need $b < 0$, or $c < 0$, or both. In either of those three cases there would be two sign changes in the polynomial, $(+-++)(++-+)(+--+)$. Hence at most two positive roots would exist. Now let's consider the negative roots. By substituting a negative value into (3.4) for I^* and considering the three cases again, only one sign change occurs for all three cases, $(---+)(-+++)(--++)$. Thus, there is exactly one negative root independent of $b < 0, c < 0$, or both. Now, since we want real roots to (3.4), this implies there are two positive roots and one negative.

Now we can consider $b = 0$ or $c = 0$. First note, if $b = 0$ then

$$N = \frac{\omega_2 \omega_4 \beta_1 + \beta_3 \omega_1 \omega_3 + \beta_3 \omega_2 \omega_3}{\omega_3 \omega_4 (\omega_1 + \omega_2)}.$$

If $c = 0$ then

$$N = \frac{\beta_3 (\beta_1 \omega_2 + \beta_2 \omega_1 + \beta_2 \omega_2)}{\omega_2 \omega_4 \beta_1}.$$

Thus, we cannot have $b = c = 0$.

First we will first consider $b = 0$. In this case (3.4), reduces to

$$aI^{*3} + cI^* + d \tag{3.5}$$

Again applying Descartes' Rule of Signs we see if $c > 0$, then there are no sign changes, $(+ + +)$. Thus, $c < 0$ is needed for biological feasibility and hence there are two sign changes, $(+ - +)$. Therefore, there are at most two positive roots. Now we will look at the negative roots. Substituting a negative number into (3.5) gives the signs $(- + +)$, so there is one sign change and thus one negative root. As above we want all the roots to be real, so there are two positive roots and one negative root.

Now we will do the same for $c = 0$. Making this substitution into (3.4) we get,

$$aI^{*3} + bI^{*2} + d \tag{3.6}$$

By Descartes' Rule of Signs we see if $b > 0$, then there are no sign changes, $(+ + +)$, and so we require $b < 0$. If $b < 0$ then we have two sign changes, $(+ - +)$. Substituting a negative value into (3.6) we only have one sign change, $(- - +)$. Therefore, since we want real roots to (3.6) we again have two positive roots and one negative. Thus, to have a biologically feasible endemic equilibrium we need $b \leq 0$, $c \leq 0$, or $b < 0, c < 0$. In all of these instances we see our endemic equilibrium may not be unique.

In order to determine the stability of the equilibrium points for the Three Exposed Class Model we will use linearization. Based on our work with the One Exposed Class Model and the similarity in form the equilibrium points from both models have, we expect that linearization will fail to determine stability due to non-isolated fixed points. The Jacobian

for (3.2) is,

$J =$

$$\begin{bmatrix} -I\mu_1 - I\mu_2 - I\mu_3 & 0 & 0 & 0 & -\mu_1 S - \mu_2 S - S\mu_3 \\ I\mu_1 & -I\omega_1 - I\omega_2 & \beta_1 & 0 & -\omega_1 E_1 - \omega_2 E_1 + \mu_1 S \\ I\mu_2 & I\omega_1 & -I\omega_3 - \beta_1 & \beta_2 & \omega_1 E_1 - \omega_3 E_2 + \mu_2 S \\ I\mu_3 & I\omega_2 & I\omega_3 & -I\omega_4 - \beta_2 & \omega_2 E_1 + \omega_3 E_2 - \omega_4 E_3 + S\mu_3 + \beta_3 \\ 0 & 0 & 0 & I\omega_4 & \omega_4 E_3 - \beta_3 \end{bmatrix}.$$

For VFE/SVFE we get,

$$J|_{(S^*, N-S^*, 0, 0, 0)} = \begin{bmatrix} 0 & 0 & 0 & 0 & -\mu_1 S^* - S^* \mu_2 - S^* \mu_3 \\ 0 & 0 & \beta_1 & 0 & -(N - S^*) \omega_1 - (N - S^*) \omega_2 + \mu_1 S^* \\ 0 & 0 & -\beta_1 & \beta_2 & (N - S^*) \omega_1 + S^* \mu_2 \\ 0 & 0 & 0 & -\beta_2 & (N - S^*) \omega_2 + S^* \mu_3 + \beta_3 \\ 0 & 0 & 0 & 0 & -\beta_3 \end{bmatrix}.$$

Note, since this matrix is upper triangle that $\lambda_1 = \lambda_2 = 0, \lambda_3 = -\beta_1, \lambda_4 = -\beta_2, \lambda_5 = -\beta_3$.

Thus, due to the 0 eigenvalue stability cannot be determined for the VFE/SVFE.

Next we linearize around the endemic equilibrium (EE). Observe,

$J|_{EE} =$

$$\begin{bmatrix} -I^*(\mu_1 + \mu_2 + \mu_3) & 0 & 0 & 0 & 0 \\ V\mu_1 & -I^*(\omega_1 + \omega_2) & \beta_1 & 0 & -\frac{\beta_3 \beta_1 \beta_2 (\omega_1 + \omega_2)}{(\omega_3 (\omega_1 + \omega_2) I^* + \beta_1 \omega_2) I^* \omega_4} \\ I^* \mu_2 & I^* \omega_1 & -I^* \omega_3 - \beta_1 & \beta_2 & -\frac{\beta_3 (\omega_3 (\omega_1 + \omega_2) I^* - \beta_1 \omega_1) \beta_2}{(\omega_3 (\omega_1 + \omega_2) I^* + \beta_1 \omega_2) I^* \omega_4} \\ I^* \mu_3 & I^* \omega_2 & I^* \omega_3 & -I^* \omega_4 - \beta_2 & \frac{\beta_3 \beta_2}{I^* \omega_4} \\ 0 & 0 & 0 & I^* \omega_4 & 0 \end{bmatrix}.$$

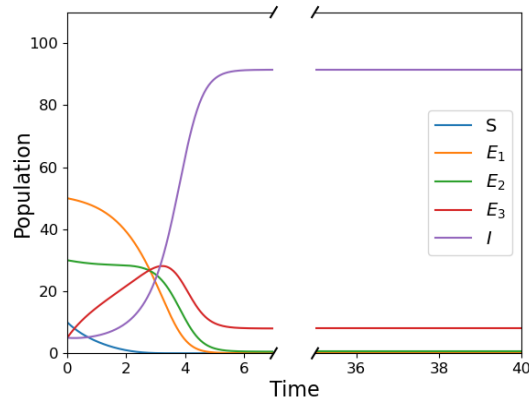
Since we can't determine the eigenvalues directly from the matrix we must compute the characteristic equation. Using Maple and simplifying we get the characteristic equation is,

$$\begin{aligned}
P(\lambda) = & -\frac{1}{\omega_3(\omega_1 + \omega_2)I^* + \beta_1\omega_2}(I^*(\mu_1 + \mu_2 + \mu_3) + \lambda)(\omega_3^2\omega_4(\omega_1 + \omega_2)^2I^{*4} \\
& + (\omega_1 + \omega_2)((\omega_3 + \omega_4)\omega_2 + (\omega_4 + \omega_1)\omega_3 + \omega_4\omega_1)\lambda + 2\omega_2\omega_4\beta_1)\omega_3I^{*3} \\
& + (\omega_3(\omega_1 + \omega_2)(\omega_2 + \omega_3 + \omega_4 + \omega_1)\lambda^2 + ((2\beta_1 + \beta_2)\omega_3 + \omega_4\beta_1)\omega_2^2 \\
& + ((2\omega_4 + 2\omega_1\beta_1 + 2\beta_2\omega_1)\omega_3 + \omega_4\beta_1\omega)\omega_2 + \omega_1\omega_3(\omega_4\beta_1 + \beta_2\omega_1))\lambda \\
& + (\beta_1^2\omega_4 - \beta_3\beta_2\omega_3)\omega_2^2 - 2\omega_2\omega_3\beta_2\beta_3\omega_1 - \omega_3\beta_2\beta_3\omega_1^2)I^{*2} + (\omega_3(\omega_1 + \omega_2)\lambda^3 \\
& + (\omega_2^2\beta_1 + ((2\beta_1 + \beta_2)\omega_3 + \beta_1(\omega_4 + \omega_1))\omega_2 + \omega_1\omega_3(\beta_1 + \beta_2))\lambda^2 \\
& + (\beta_1(\beta_1 + \beta_2)\omega_2^2 + (\beta_2(\beta_1 - \beta_3)\omega_3 + \beta_1^2\omega_4 + \beta_1\beta_2\omega_1)\omega_2 + \omega_3\beta_2\omega_1(\beta_1 - \beta_3)\lambda \\
& - 2\omega_3\beta_1\beta_2\beta_3(\omega_1 + \omega_2))I^* - \beta_1\omega_2(\beta_2\beta_3 - \beta_2\lambda - \lambda^2)(\beta_1 + \lambda))\lambda.
\end{aligned}$$

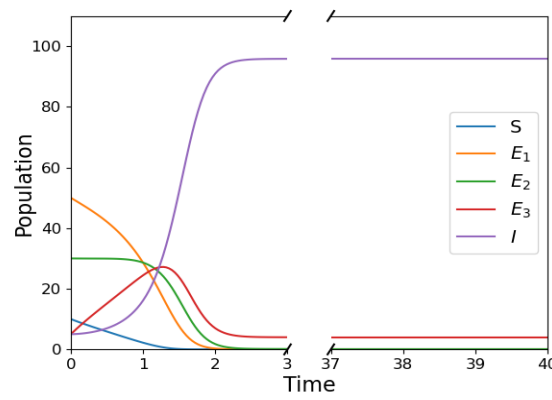
This is a fifth order polynomial in λ and due to it's complexity Maple is unable to compute it's roots. Although, we note that since $P(0) = 0$, there must be at least one eigenvalue that is zero. Thus, linearization is not able to determine stability for (3.2), so the dynamics of Three Exposed Class Model are explored further with numerical simulations. While attempts where made to reduce the system, in a similar manner as done with the One Exposed Class Model, due to the complexity it could not be analyzed in way that produced useful information.

3.3.2 Numerical Simulations

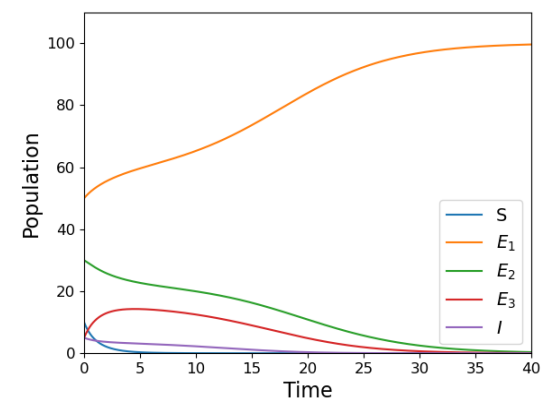
Figure 3.5a shows model simulation results in the example population using parameter values chosen as the base case scenario. Subsequent simulation results exploring parameter modifications are then compared back to this base case. We assume there is only a low level of protective factors and intervention/disruption techniques being employed in the base case scenario. Figure 3.5a shows that in the long term, the Susceptible and Exposed₁ classes approach a population of zero, the Exposed₂ population approaches 1 individual, the Exposed₃ population approaches 8 individuals, and the Infectious class approaches a population of 92. It is noted that the transient dynamics resolve rather quickly (by approximately $t = 5$), before nearing the steady state value.



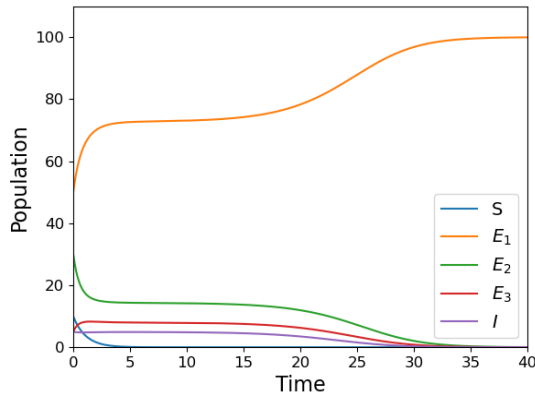
(a) Base case parameter combination



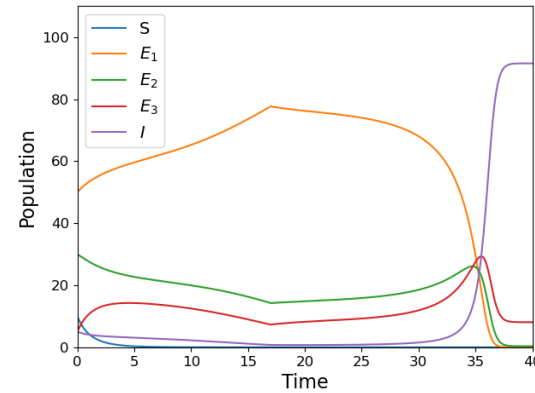
(b) Increased propensity for violence after exposure



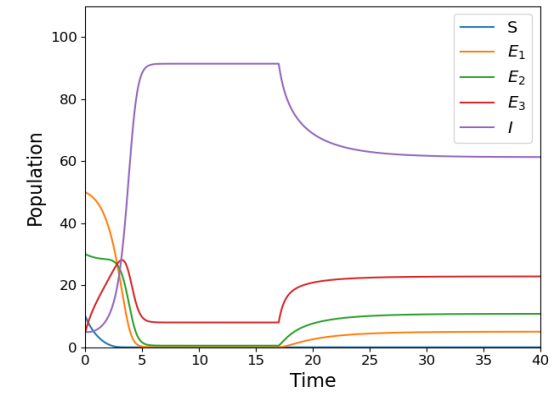
(c) Intervention measures increased



(d) Increased support after exposure



(e) Intervention measures decreased



(f) Increased intervention measures introduced

Figure 3.5: Numerical Simulations. Example simulations of Three Exposed Class Model with varying parameter values. (a): Illustrates a community with a low level of protective factors and intervention/disruption techniques (b): Shows an increase in the population's propensity to commit violence after exposure (c): Increases the amount of intervention measures that decrease propensity for violence after exposure (d): Increases the amount of intervention measures that promote healing after an increase of violence propensity (e): Starts with the same parameters as (c) and then intervention measures are decreased (f): Starts with the same parameters as (a) and then intervention measures are increased.

Figure 3.5b shows the results of a simulation where the ω_i parameters are doubled in value from that used in the base case scenario and all other parameters remain at their base case values. This increase in the population's propensity to commit violence after exposure results in the long-term population values of the Susceptible, Exposed₁, and Exposed₂ to approach zero, of the Exposed₃ to approach a population of 4, and of the Infectious class to approach a population of 96. It is also noted that the transient dynamics approach to the steady state in less than half the amount of time it took the base case simulation.

While Figures 3.5c and 3.5d have similar long-term behavior, their transient behavior is quite different. In Figure 3.5c, the ω_i parameters are decreased by half of the values used in the base case, while the other parameters stay at the base case values; this is the opposite relationship that resulted in the time courses in Figure 3.5b. In either of these scenarios (Figure 3.5c and Figure 3.5d), the entire population is transferring to the Exposed₁ class and has largely done so by $t = 40$. In Figure 3.5d, a similar behavior can be seen by instead varying $\beta_{1,2}$, and leaving the other parameters at baseline, including the ω_i parameters. For this scenario, β_1 and β_2 must be increased approximately five fold from their baseline values to recreate an outcome comparable to Figure 3.5d. Again it is noted that in the long-term the entire population will move into the Exposed₁ class.

Figures 3.5e and 3.5f show simulations with dynamic parameter behaviors. In the previous four figures the parameters stayed constant throughout the entire run, but in Figures 3.5e and 3.5f a parameter change is introduced at $t = 17$. Figure 3.5e started with the same parameter combination as Figure 3.5c, but at $t = 17$, ω_i are increased and $\beta_{1,2}$ are decreased. By approximately $t = 37$ this run approaches a similar steady state as those in Figures 3.5a, 3.5b. The simulation resulting in Figure 3.5f starts with the same parameters in Figure 3.5a, but at $t = 17$, $\beta_{1,2}$ are doubled and ω_i are decreased by approximately a third. Figure 3.5f approaches its steady state quicker than Figure 3.5e-around $t = 27$. In this run the transient dynamics level off with a much lower infected population (and respectively higher exposed populations) than it would have without the parameter change. While all four of these simulations are examples, we see that the behavior is qualitatively similar to observed

real world scenarios [24, 93, 8] and thus exploring these model simulations may provide useful insights into the dynamics of the system.

3.4 Discussion

3.4.1 Equilibrium and Stability Analysis

By exploring the real world interpretation of the equilibrium points we find our models are in keeping with the general understanding of violence. The VFE represents a population where all people exposed to and committing violence are removed, thus the remaining population would be all susceptible. While this would be ideal, as there would be no effects of violence on the population, it most likely will never occur in a real world situation. This is reflected in both models as the VFE is not stable. The SVFE is similar to VFE, except instead of having a totally susceptible population a portion is exposed. Although, even with some exposure in the population no one is committing any new acts of violence. This equilibrium could be seen as the complete success of intervention and disruption techniques and protective factors in a community, as no one's propensity for violence ever passes the threshold for taking action. This outcome could be more realistic than the VFE, although still extremely unlikely in actual communities as it requires the end of violence. The endemic equilibrium represents when violence is continually occurring in the population and, thus, people are continually being exposed to it. This outcome is most realistic and what is reflected in the world. Since some level of violence has always been present in communities, in the long term we should see the population primarily divided among those with exposure and those committing violence. This also accounts for a more realistic implementation and response to the intervention and disruption techniques and protective factors where they do not have a perfect success rate. We note that this is reflected in the One Exposed Class Model's stability of the endemic equilibrium. Thus, the interpretation from our equilibrium analysis show that applying a mathematical epidemiological model to violence leads to realistic equilibrium points that can be seen in the literature and in reality.

Since it is unlikely to completely eradicate violence, our goal is to reduce the population of people actively committing violence with the limited resources available. We can explore different ways of accomplishing this through our mathematical models by determining the optimal parameter combination that leads to reduced amounts of violence in the long term. For the One Exposed Class Model we want to find the optimal way to reduce $\frac{N\omega - \beta}{\omega}$, which would lower the number of infectious individuals at equilibrium. This could be done by forcing ω to be small relative to β . While this would maximize the number of people exposed, it would minimize the number of infectious individuals, which is the ideal outcome for the endemic equilibrium. This optimization would not be straight forwards as it has to take into consideration the constraints present in real life scenarios, such as funding limitations for protective factors and political climates in communities. On the other hand, the Three Exposed Class Model has two, potentially unique, endemic equilibrium points. Thus, if these equilibrium points are unique we can study what parameter combinations would lead to the equilibrium point with the lower amount of violence in the long term. We explore this idea further with our numerical simulations. By analyzing the equilibrium points and stability of the models we are able to learn about the real world system without needing multiple empirical studies which cuts down on time, resources, cost, and need for new data.

3.4.2 Numerical Simulations

The simulations provide an illustration that shows how the Three Exposed Class Model can be useful in studying real world situations, especially as appropriate data is collected and integrated. For example, our simulations show that with low levels of protective factors/interruption and disruption techniques violence spreads rapidly through a community, such as what is seen in Figures 3.5a and 3.5b. This result reflects what may happen in extreme situations when communities are struggling with high rates of violence and have weak community support systems. Further, these simulations also provide a useful strategy for the most effective way to handle violence in a community. Considering the implications of Figures 3.5c and 3.5d, we see that investing in violence intervention measures is much more effective then investing in support after exposure. By decreasing in half the impact

of violence exposure, the amount of violence in a community goes to zero in the long-term; but to reach this same long-term outcome through an increase in the amount of healing after an exposure, the increase must be approximately 5-fold. This implies that investing in programs that give individuals the capacity to cope with violence exposure in a healthy manner (protective factors) are more effective at preventing the spread of violence than programs focused on decreasing a person's propensity to commit violence once that propensity has been increased due to exposure (interruption/disruption techniques). This is, of course, a result seen in epidemiological models: in stopping epidemic outbreaks, it is more effective to prevent people from catching a disease, than treating it once someone is sick.

Further, Figures 3.5e and 3.5f illustrate what can happen when interventions are introduced or eliminated from a community. Figure 3.5e shows a situation where a community has good levels of intervention /disruption techniques and protective factors, then they drastically decrease those levels (for reasons such as a change in funding or priority switch by those leading a community). Although the effects of good intervention measures last for a period of time (approximately 13 time units), eventually the community will end up in the same situation as if there had never been adequate preventative measures. On the other hand, in Figure 3.5f it is observed that once violence is freely spreading within a community, it is much more difficult to get it under control. Even though introducing higher levels of intervention/disruption techniques and protective factors to the run did lower the level of those in the infectious class, there was still high amounts of violence occurring as it approached the steady state. Thus, based on these examples, the negative effects of violence exposure seem to have a greater impact in a community than the positive effects of preventative measures. So, these measures need to be introduced to a community early and consistently to maximize their effects.

In general, these insights are in keeping with the qualitative understanding of violence. Empirical case studies such as [24, 93] look at the effectiveness of Cease Fire-like organizations in violence reduction. Statistical analysis of real-world data allowed both papers to conclude that using community-based public health prevention measures help to reduce

future occurrences of violence. Our model is in qualitative agreement with these data-backed findings that show increasing preventative measures decrease the amount of violence in the communities, as seen in Figure 3.5c and Figure 3.5f. Further, studies like [8] look at effectiveness of implementing interventions for those who are hospitalized due to violence. This study shows youths who received these interventions are less likely to commit violence themselves after recovery than those who do not, just as Figure 3.5d and 3.5f predict. Since our models agree qualitatively with data and results from the field, this suggests that the quantitative formulation we have constructed is appropriate for the system.

3.4.3 Public Health Policy

Due to the following well-understood dynamics of violence exposure, we are able to successfully apply a mathematical epidemiological framework and consider violence from a public health perspective: exposure to violence increases the likelihood a person will be further exposed to violence [64, 34], exposure to violence, especially repeated exposures [64], in adolescence increases an adolescent’s propensity to commit violence [47, 36], and protective factors, interruption/disruption techniques, and the passage of time can help decrease a person’s propensity to commit violence after it has been increased due to exposure [93, 39, 36, 24, 8]. As these dynamics closely resemble that of infectious disease spread, our proposed epidemic model for violence is well-grounded. Note, our model does not attempt to tackle the root causes of violence, just as mathematical epidemic models do not propose direct methods to combat germs or how they cause disease, but instead focus on how to limit their spread and impact. These epidemiological tools are still widely accepted and critical tools in the understanding of outbreak dynamics. Thus, like mathematical epidemic models, our violence model still adds a useful quantitative perspective for prevention and mitigation techniques to the literature.

Having valid quantitative models that capture qualitative understanding allows us to make quantitative predictions for relative efficacy of potential intervention strategies. For those working in the field, like social workers and public health leaders, tools like these can provide

results and insight to help determine where effort may be most effective, especially when working with limited resources. For example, the implicit time scale of our model, which focuses on adolescents, aged 14-17, will allow us to explore the most effective age targeted interventions. This could have an immediate impact on the public health of communities by reducing the frequency of, and damage caused by, violence. Further, these types of results can also help inform policy makers and community advocates. Mathematical modeling can provide a more time-and-cost-effective approach to gain insight and results to help with allocating and securing funding for violence intervention measures. It can also guide public health policy, law enforcement, and community leaders on the most effective ways to handle violence in their communities. This can be especially important if the most effective ways are found to be different from those in current use. For example, our simulations suggest the most effective way to make use of community funding is investing in protective factors, like strengthening economic support for families, changing social norms surrounding violence, providing access to quality low-cost childcare, and offering after school/mentorship programs [36]. Having a general framework in place to study complex system such as those proposed in this work will allow us in the future to customize the model to represent specific communities or types of violence. This view of violence as a multi-scale dynamical system offers a new perspective and mechanism to help stem the epidemic of violence.

3.4.4 Limitations of the Model

Like all mathematical models and simulations, the ones presented here have limitations. The simulations presented in this chapter are illustrative in nature as we have not yet incorporated quantitative data from the field. Thus, while the example simulations illustrate the approach and provide a glimpse of how this system works and responds, we are not suggesting that this model provides an accurate representation of what happens in actual cities. Further, the assumptions mentioned earlier also place limits on the model. These assumption stem from decreasing model complexity, especially in the absence of data. In the future, we plan to address model limitations. By introducing more complexity to the model, some of the limitations could be addressed. One way this can be accomplished is by including additional parameters in the model(s) to explore other features and characteristics

of violence, such as introducing spontaneous acts of violence into the community (violence with causes other than exposure, such as mental health or physical illnesses) [33] or explicitly exploring the differences in types of violence exposure and who is most at risk [34]. Although, careful consideration would have to be taken when increasing complexity so that useful analysis can still be performed and interpreted. Including data is another way to address some of these limitations. Due to the sensitive nature of this work, we acknowledge that accessing the needed data may be difficult, but we aim to acquire and integrate data on violence from sources such as the National Violent Death Reporting System, the National Crime Victimization Survey, and FBI’s Uniform Crime Reports. This will allow for better construction of model functional relationships as well as the ability to better estimate model parameters. Of course, special consideration will need to be taken to account for any implicit biases that may be embedded in the data. With the inclusion of carefully interpreted data, the models developed will provide more accurate descriptions of the real-world system and thus better predictions for real communities.

3.5 Basic Reproductive Number

An important topic in traditional disease modeling is the basic reproductive number, called $\mathcal{R}_0$. In epidemiology it is defined as the typical number of secondary infections caused by one infectious individual in a susceptible population, while the traditional mathematical definition is the parameter threshold that determines if $\dot{I} > 0$ or $\dot{I} < 0$. Based on these definitions we get the condition that if $\mathcal{R}_0 < 1$ then the disease will die out and if $\mathcal{R}_0 > 1$ then the disease becomes endemic. While the models proposed in this chapter are based on concepts from traditional disease modeling, the basic reproductive number is an aspect where the models diverge from the analogy. For instance, unlike with infectious diseases, exposure to a single violent act by an individual does not directly cause another individual to commit violence, in most cases. Whereas, exposure to an individual with an infectious disease will directly cause transmission with some probability. For violence, repeated exposure may be needed to increase a person’s propensity to commit violence and that eventually leads to a violent act being perpetrated. Further, though not explicitly explored within these models,

there is always the potential for novel emergence of violence from an unexposed individual, which is not possible for infections. A more specific issue with the One Exposed Class Model is if the endemic equilibrium exists then it should be stable, because the biological feasibility condition is equivalent to $\mathcal{R}_0 > 1$. This isn't ideal when studying diseases or, in our case, violence as it means there is not even a theoretical opportunity for violence to cease once it is introduced to the population. Further, for the Three Exposed Class Model traditional methods for calculating $\mathcal{R}_0$, like the Next Generation Matrix method, are not applicable due to the possible non-uniqueness of the endemic equilibrium. Thus, the traditional concept of the basic reproductive number and its way to calculate it does not translate well to applications outside of disease modeling.

For these reasons, in the context of violence it will be useful to re-imagine the concept of the basic reproductive number in a way that isn't so strictly tied to disease modeling and stability of equilibrium points. This more general framework could then work in many different applications. Thus, the idea of the contributive number, $\mathcal{C}_0$, is proposed. For violence, $\mathcal{C}_0$ could describe the damage an act of violence contributes to the collective mental health of a generation. Note, due to the potential for novel violent acts in the absence of exposure, it may be that the contributive number can remain below 1, while violence increases in a population. Understanding this contribution of exposure relative to other factors is part of what such a model can help to characterize and clarify.

3.5.1 Contributive Number

Since we want $\mathcal{C}_0$ to capture the amount of damage done to a community, we can formulate the system into a flow network which can then model the flow of damage (due to violence) through a community using the techniques introduced in Chapter 2. While the traditional compartmental model tracks the flow of people through classes, the flow network model will instead track the flow of damage through nodes. Recall that Chapter 2 provides the technical details of the process of converting a ordinary differential equation compartment model to a flow network and then using it to calculate the basic reproductive number. We will apply

this technique to our Three Exposed Class Model.

Figure 3.6 represents how damage from violence flows through the system for the Three Exposed Class Model. In the digraph we see that node I is the source and node E_3 is the sink. Since damage is caused by exposure to violence, then clearly the damage is originating in the infectious class/node. E_3 is the sink as it acts as a bottleneck for flow through the system. Note that in order for the flow of damage to return to I , all of the flow must first go through E_3 . Thus, E_3 is the sink for the flow of damage due to violence. Now adding the edge weights that come from (3.2) we get the full network in Figure 3.7.

Next we evaluate the network at the VFE. For disease modeling applications this means that the total population is in the susceptible, S , class. In our violence model, this idea of disease free isn't applicable as there are no communities that are completely free from the effects of violence. Thus, we consider the VFE to be the entire population being in the exposed classes. So, we assume that $S = 0, E_1 + E_2 + E_3 \approx 1, I^* > 0$ (but small). The new edge capacities are shown in Figure 3.8.

We can then reduce the network model by eliminating the edges with zero capacity and the isolated node (S) as seen in Figure 3.9. Now we solve the minimum cut problem by finding the cut sets, seen in Figure 3.10a-3.10d. We see that Figure 3.10d has the cut with the minimum edge weight of $\omega_2 E_1^* I^* + (\omega_3 I^* - \beta_1) E_2^* + (\omega_4 I^* - \beta_2) E_3^*$. With the cut sets defined, let's consider their biological feasibility. Figure 3.10a isolates the infectious class from the rest of the population. Thus, this cut set insulates violence exposure from spreading. This would look like isolating anyone who is committing violence away from the population who is not committing violence. While this is biologically feasible, it is not realistic. Figure 3.10b focuses all the the resources on the E_1 class, so they do not progress to further exposed classes. This would prevent people in the community from ever having enough exposure to cause them to commit violence.

Figure 3.10c represents an approach to violence prevention where resources are concentrated

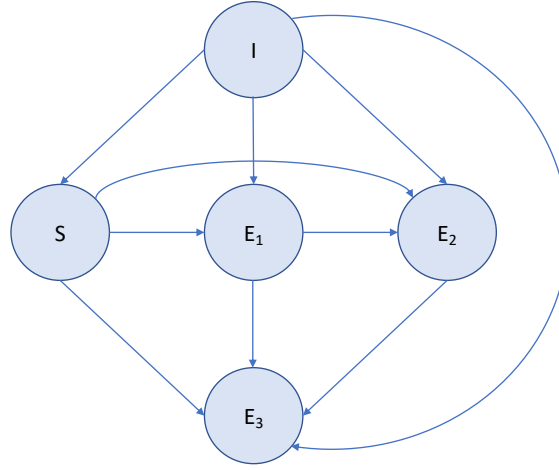


Figure 3.6: Unweighted Flow Network. Flow network digraph for how damage flows through a community in the Three Exposed Class Model. See Chapter 2 for more details on how it was constructed.

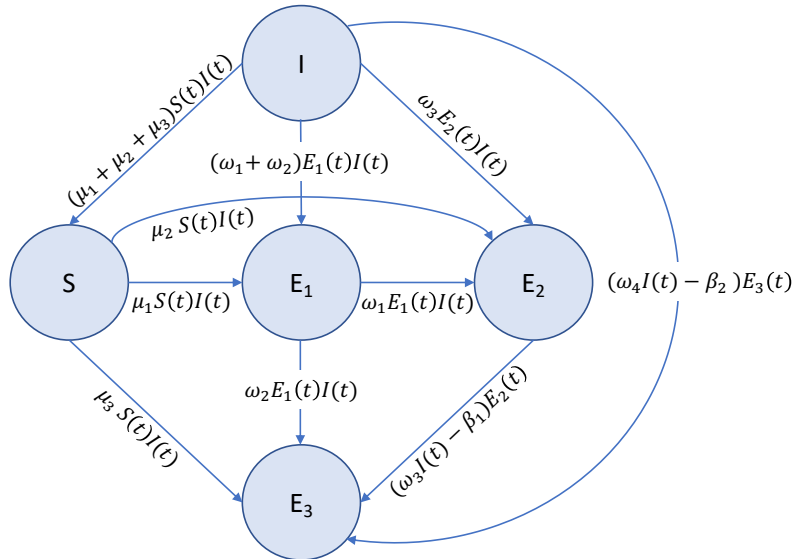


Figure 3.7: Weighted Flow Network. Flow network with dynamic edge added

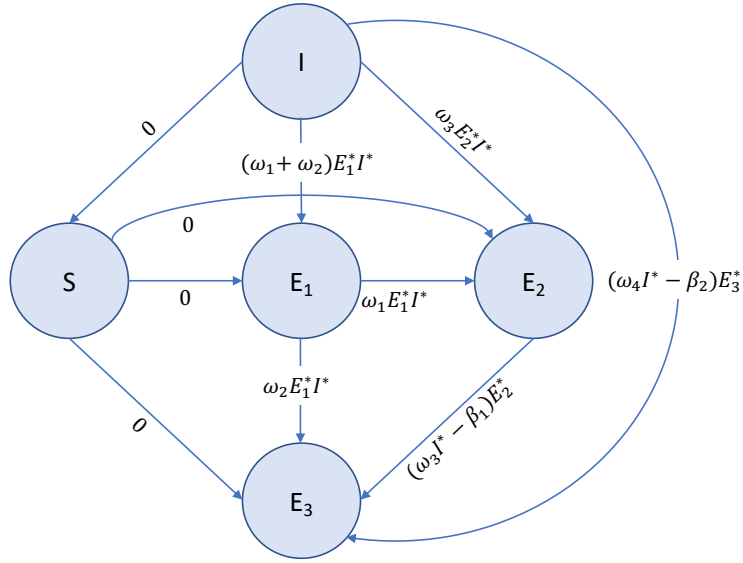


Figure 3.8: Flow Network Evaluated at VFE.

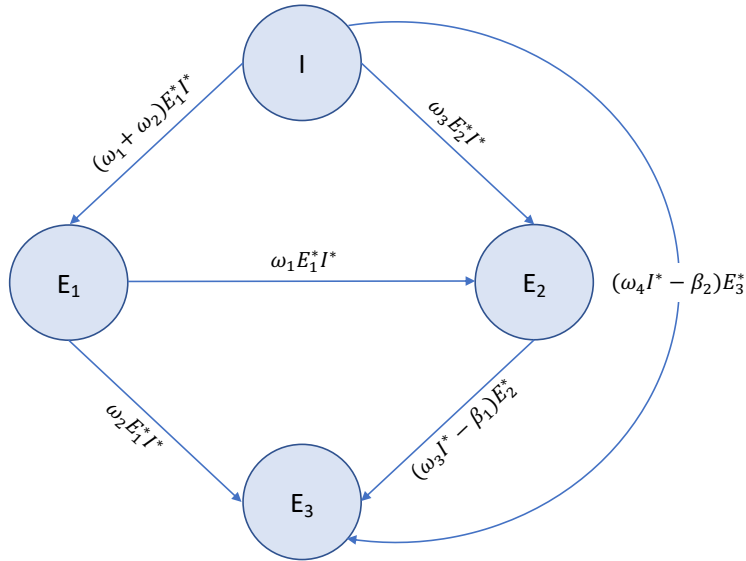
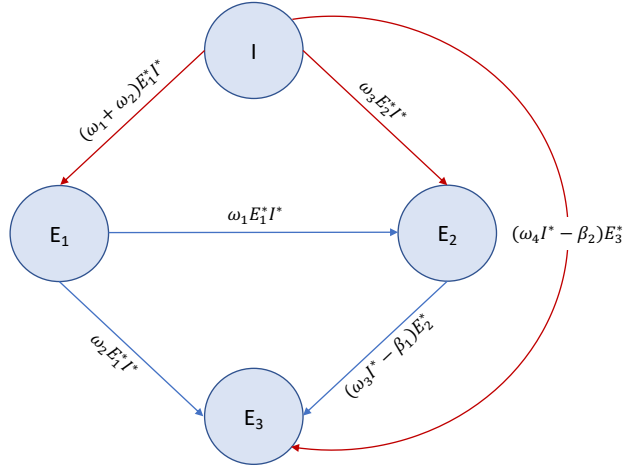
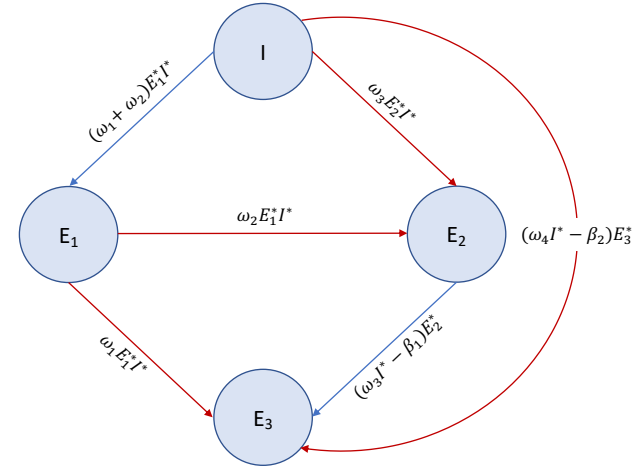


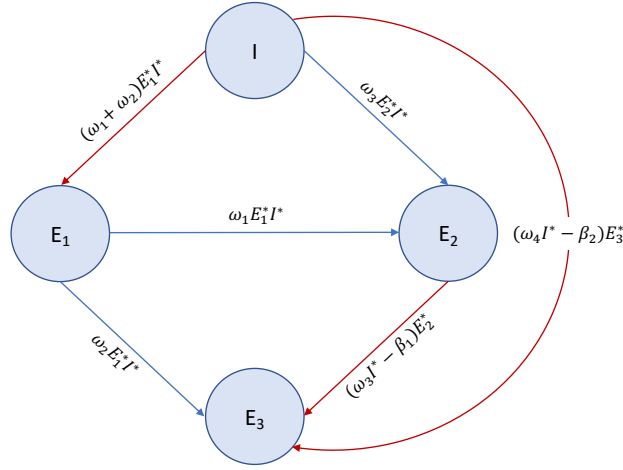
Figure 3.9: Reduced Flow Network



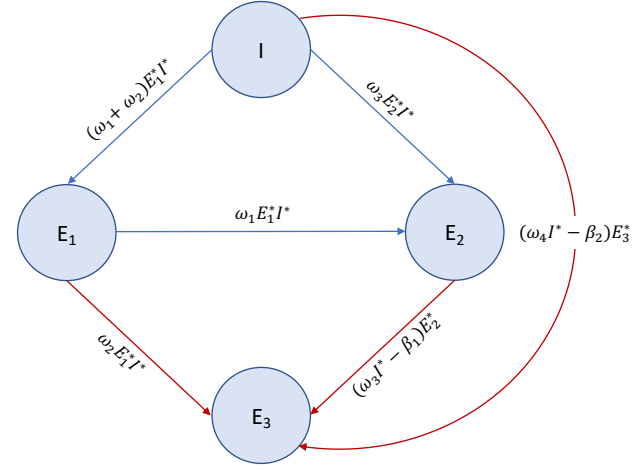
(a) Total weight: $(\omega_1 + \omega_2)E_1^*I^* + \omega_3E_2^*I^* + (\omega_4I^* - \beta_2)E_3^*$



(b) Total weight: $(\omega_1 + \omega_2)E_1^*I^* + \omega_3E_2^*I^* + (\omega_4I^* - \beta_2)E_3^*$



(c) Total weight: $(\omega_1 + \omega_2)E_1^*I^* + (\omega_3I^* - \beta_1)E_2^* + (\omega_4I^* - \beta_2)E_3^*$



(d) Total weight: $\omega_2E_1^*I^* + (\omega_3I^* - \beta_1)E_2^* + (\omega_4I^* - \beta_2)E_3^*$

Figure 3.10: Cut Sets for Flow Network. Visual representation of the cut sets for 3.9, where the edges colored red represent the cut edges. The total weight of the cut (red) edges is provides beneath each digraph.

on high risk groups. This is so they don't go on to expose their community even though they are receiving exposure from both I and E_1 . An example of this could be a high level of resources and support focused at an at risk high school, while lower risk schools don't receive the same support. This approach would have to emphasize support after exposure, to stop violence from continuing. Finally, Figure 3.10d, which is a our minimum cut, represents the most realistic scenario. In this case people are being exposed but there are enough resources available to prevent them from committing violence themselves. Thus, with all the support and interventions available to everyone violence would cease. Since this is the minimum cut, by the Max Flow Min Cut Theorem it is the max flow and so $\mathcal{C}_0 = \omega_2 E_1^* I^* + (\omega_3 I^* - \beta_1) E_2^* + (\omega_4 I^* - \beta_2) E_3^*$ for the Three Exposed Class Model. Note we do not need to solve $\mathcal{C}_0 > 1$ as this is not the traditional basic reproductive number and so we are only interested in the max flow.

3.6 Conclusion

As violence continues to wreck havoc in our communities, it is clear that viewing violence solely as an issue for the criminal justice system is not effective at preventing the cycle of violence. The promising approach of implementing community-based intervention models to understand and prevent violence has the potential to save countless lives and avoid unnecessary physical and mental trauma. By expanding on the work done by physicians and public health experts, the use of Immuno-Epidemiological modeling can be a valuable additional tool to help overcome this societal challenge. In this chapter, we formalize the analogy between violence and infectious disease, which expands the groundwork needed to apply the mathematical tool of epidemiology. Then the One and Three Exposed Class Models for violence spread are formulated and the potential usefulness of the model is explored by equilibrium and stability analysis and running some example simulations. Although, these models are heuristically defined rather than calibrated to data, we nevertheless gain preliminary insights into how different strategies for preventing the spread of violence differed in effectiveness. By comparing the outcomes of our simulations to results from empirical studies, we confirm this approach is an appropriate way to model this system. Further, we

developed the idea of the contributive number that can be applied not only to the models proposed here, but for other social modeling applications. By following the format of the Immuno-Epidemiological modeling structure, once the within host model is integrated, we will be able to study and gain insight into the bidirectional feedback between individuals and their communities when exposed to violence. In conclusion, we see that a public health approach to violence will be an effective way to help curb this epidemic and that mathematical modeling can be an important and useful tool in achieving this goal.

Chapter 4

Connecting People to Food: A Network Approach to Alleviating Food Deserts

Publication information: In review

Disclosure statement: Anna Sisk is the first and corresponding author on the manuscript based on this chapter. Her co-authors are Nina Fefferman (advisor), Kristen Rappazzo, and Tom Luben. Sisk and Fefferman conceived the work. All the authors contributed to the understanding of the problem and formulation of the analysis. Sisk implemented the analysis and drafted the manuscript. All the authors worked to revise, review, and approve the manuscript. Sisk received assistance with the software used for the analysis from Mona Papeş and Monica Jimenez.

Abstract

In 2020, 13.8 million people in the United States struggled with food insecurity, meaning they were uncertain about their ability to ensure their food needs would be met. Whether a person struggles with food security can be influenced by where they live. Food deserts are census tracts that experience both high rates of poverty (at least 20 percent of residents living at or below federal poverty thresholds) and low access to grocery stores that offer fresh and nutritious foods. Food deserts and food insecurity disproportionately affect minority and disadvantaged communities and may contribute to serious health issues like diabetes, high blood pressure, and obesity. Public policies can be utilized to lessen the impact of food deserts and one mechanism city leaders can employ to achieve this is their public transit system. We characterized the role public transportation plays in connecting food desert residents with food by formulating network models from data on food deserts, grocery stores, and public transportation systems for five representative locations: Brown Deer, WI; Lawrence, KS; Albuquerque, NM; Charlotte, NC, and Raleigh, NC. We analyzed these networks by looking at centrality measures, specifically degree and closeness. These centrality measures provide insight on the current situation regarding grocery store access for food deserts. We found that not all food deserts have the same levels of connectivity-based limitations in access. Thus, using these results, we identify how public transportation could

better connect people with nutritious food and offer these suggestions to city leaders as a way to help eradicate food deserts.

4.1 Introduction

In 2020, 13.8 million people in the United States struggled with food security, meaning they were unable to ensure or uncertain that their food needs would be met [17]. While struggling with food security has negative physical and psychological impacts on people of all ages, such as increased mental health issues and poor physical health outcomes, children face the most severe consequences [45]. Children living or born into food insecure homes are at an increased risk of birth defects, anemia, cognitive issues, aggression, anxiety, asthma, depression, suicide ideation, and overall poorer physical and mental health outcomes than their food secure counterparts [45]. Whether a person struggles with food security can be influenced by where they live. Food deserts are census tracts that are both low income and low access, where low income is defined by the United States Department of Agriculture (USDA) to have a poverty rate of 20 percent or greater, or a median family income at or below 80 percent of the statewide or metropolitan area median family income and low access is defined to be at least 500 persons and/or at least 33 percent of the population living more than 1 mile from a supermarket or large grocery store [68]. While the existence of food deserts does not always imply its inhabitants are food insecure, it is often a contributing factor. Further, whether or not a person is food insecure, living in a food desert is associated with increases in the likelihood of health issues such as diabetes, high blood pressure, high cholesterol and obesity [62] due to decreased access of healthy and nutritious foods. Thus, addressing and eliminating food deserts is an important public health and social welfare objective.

While the ideal solution to ending food deserts is to increase the number of grocery stores near food deserts and/or lift communities out of poverty, this is rarely a realistic option. According to [30] grocery stores and supermarkets are unlikely to be voluntarily located in food deserts due to economic reasons such as, higher urban land, labor, and utility costs,

low profit margins on more perishable food items, and increased theft problems. Since city leaders and local governments do not have control over where grocery stores are located and it can be economically prohibitive to incentivize them to open in low income and urban areas, they must come up with alternative solutions to eliminate food deserts. Some of the most effective ways to do this are through policy interventions. Strategies such as revisiting city zoning codes to allow for grocery stores to open in previously unavailable areas [26], streamlining and prioritizing food licenses [26], and providing financial backing for fresh food initiatives [57] can help reduce the amount and effects of food deserts. While city governments cannot always bring grocery stores to people, they can help bring people to grocery stores. Offering shuttles that can transport food desert residents directly to and from grocery stores and working with rideshare and taxi companies to offer reduced fares to and from grocery stores can help increase access to food [26]. In some urban and suburban areas, the public transit system can play an important role in connecting people to food. Since food deserts, by definition, are low-income communities, public transportation is often the most cost-effective way for people to travel around their cities. In addition, city leaders and departments have more control over their transit systems than they do the location of grocery stores. Thus, analyzing the public transit system can provide useful information for local decision makers working to eradicate food deserts.

Mathematical modeling is a useful tool for analyzing both public transit and food deserts. Agent-based models of food deserts and food access have allowed for the assessment of different interventions to increase food access, including adjustments to the public transit system [95, 3]. Network models are commonly used for studying public transit [25, 38, 77]. A network, sometimes referred to as a graph, is a collection of nodes and edges, where the nodes are connected to each other by the edges. This framework lends itself easily to public transportation systems as transit stops are connected to each other by transit lines. Thus, using a network model is both a time- and cost-effective way to study the relationship between public transit and food deserts. The results of such network models can provide valuable information to local decision makers working to determine how public transit can

be better utilized to help alleviate issues stemming from food deserts.

In this chapter, we characterize the role public transportation plays in connecting food desert residents with food by formulating network models from data on food deserts, grocery stores, and public transportation systems, where the public transportation lines serve as the edges and the food deserts and grocery stores are the nodes. We then analyze the networks by performing centrality measure analysis, specifically, degree and closeness. Finally, we offer some suggestions on how city leaders can use these results to help relieve problems caused by food deserts.

4.2 Methods

4.2.1 Model Formation

We identified urban and suburban locations as potential study sites and then used USDA's Food Access Research Atlas [89] to construct a preliminary list of cities/towns which contained food deserts. To simplify the analysis, and for consistency across study sites, we decided to only use sites where the public transportation system is mainly a bus system. In our final selection, we wanted to ensure heterogeneity in the study sites, so we selected both urban and suburban locations that were distributed geographically across the country. The locations we studied are Brown Deer, WI; Lawrence, KS; Albuquerque, NM; Charlotte, NC; and Raleigh, NC. Across the five study sites there are 61 food deserts distributed as follows: Brown Deer: 6; Lawrence: 5; Albuquerque: 15; Charlotte: 19; Raleigh: 16.

The conceptual framework for the network models is the food deserts and grocery stores are the nodes, while the transit lines are the edges. We implemented the network models and analyzed them in ArcGIS Pro. We primarily utilized the Network Analyst tool, Create Buffer tool, and Summarize Within tool. Each of the maps presented here and in Appendix A was created in ArcGIS Pro. Since some of the tools in ArcGIS Pro require a specific address or point, we chose one address per food desert to represent that food desert in the network. We

refer to these addresses as the representative addresses. Candidates for the representative addresses were selected by looking at the food desert locations on Google Maps. We selected addresses that have both high population densities, like apartment buildings or locations in the middle of a community (e.g. a neighborhood park or school), and that are near the center of the food desert. Selecting representative addresses in this way allows for a better understanding of the situations of real people living in food deserts. Further, it helps ensure the calculations and analysis are done where people actually live in that food desert. For example, some food deserts have non-residential areas and thus do not have permanent residents living in them. Hence, analysis done with a representative address from that area won't be as accurate for those living in that food desert. The street addresses for the representative addresses and the food desert they represent for each study site can be found in Appendix A. Note that the food deserts are referred to by the last 3 digit of their census tract number.

4.2.2 Data

Information and graphical representation of the food deserts were obtained from the USDA's ArcGIS REST Services Directory, specifically using the foodaccess2019 (MapServer) [1]. To represent the edges in the network model we created a public transportation network in ArcGIS Pro using General Transit Feed Specification (GTFS) public transit data and street centerline data. The GTFS data were obtained from Open Mobility Data [21] and a local city's website [2] and the street centerline data were obtained from each city's open data bases [65, 53, 4, 63, 18].

The USDA defines grocery stores to be establishments generally known as supermarkets and smaller grocery stores primarily engaged in retailing a general line of food, such as canned and frozen foods; fresh fruits and vegetables; and fresh and prepared meats, fish, and poultry. Convenience stores, with or without gasoline sales, are excluded [87]. We were unable to access the proprietary list of grocery stores used by the USDA, so we created our own grocery store list for each city using the Reference USA/Data Axle Reference Solutions database. We sorted the data by Standard Industrial Classification (SIC) codes and focused on the

5411-Grocery Stores category. We only included businesses where the codes in Table 4.1 were the primary SIC code, since many non-grocery stores (e.g., liquor stores, dollar stores) will have a secondary SIC code that labels them as grocery stores. Further, we included big box retailers or superstores (e.g., Walmart, Target) which are classified as department stores (SIC code 5311), but also include a grocery department. We also explicitly excluded SIC codes 5411-03: Convenience Stores and 5541-01: Service Stations-Gasoline and Oil from the search.

In order to verify the accuracy of the grocery store lists, we conducted a spot check on each city's list to ensure that only grocery stores were included. This check revealed many 'non-grocery' stores were included due to misclassified SIC codes. Thus, to increase the accuracy, we reviewed each entry on the lists by hand to determine if the store contributed to a positive food environment (offered fresh and nutritious foods in all the major food groups). The initial process was conducted using Google to search the name and/or address of the business and looking for evidence that the store sold fresh produce and meat, in addition to dry and/or packaged foods. This evidence could be in the form of pictures, customer reviews, social media, and/or websites. Businesses labeled as 'permanently closed' on Google were removed. Further, businesses that appeared (by their pictures or by their reviews) to be corner stores or neighborhood stores were also removed. These corner stores showed up most often in the Brown Deer list. These types of shops were not counted as grocery stores because although they may sell some grocery items, they act more as convenience stores, not full-service grocery stores.

Businesses that were unable to be verified through the initial Google search were separated into two groups: one where the addresses from the list matched the address from a Google search and one where they did not match. For the entries where the addresses on the list and the Google search did not match, we checked the addresses on Google Maps. If there was no evidence a grocery store was located at the street address on Google Maps, the entry was removed from the list. If there was evidence the businesses had either moved locations or changed names, the list was updated. For the businesses where the addresses did match, but were unable to be verified in the initial search a more detailed search was conducted

Table 4.1: SIC Codes. SIC codes included in the grocery store lists.

SIC Code	Description
5411-05	Grocers-Retail
5411-06	Markets-Kosher
5411-07	Grocers-Ethnic Foods
5411-08	Grocers-Health Foods
5411-09	Grocers-Take-Out-Foods
5311-04209	Retail Shops (Walmart)
5311-02209	Department Stores (Walmart)
5311-02148	Department Stores (Walmart SuperCenter)
5311-02200	Department Stores (Target)

and if there still wasn't enough evidence to definitively exclude a business, it remained on the list but was flagged to determine the effect of including or excluding these businesses on the results. Across all five study sites, there were initially 620 businesses labeled as grocery stores, but after the verification process the number of grocery stores was reduced to 408. Of the 212 businesses removed, 13 were labeled as permanently closed, 129 were food related businesses mislabeled as grocery stores in the database (e.g., corner stores, convenience stores, restaurants, liquor stores, etc.), and 70 were removed for miscellaneous reasons (addresses that did not have grocery stores at the location, businesses unrelated to food being mislabeled, duplicate entries, etc.). We conducted a sensitivity analysis between the initial grocery store lists and the lists after the verification to determine the effects of the inaccurate lists on our analysis.

4.2.3 Centrality Measures

We analyzed the networks by looking at centrality measures, which quantify specific characteristics of the contribution to, and relative importance of, each component in the overall network organization. Centrality measures can focus on nodes or edges, and which is considered to be the most 'central' is defined by the measure. There are many types of centrality measures, common ones being degree, betweenness, closeness, and eigenvalue [92]. Each centrality measure evaluates the the importance of different aspects of a network in different ways. For our network models, we considered two types of centrality measures: degree and closeness. Both of these measures focus on which node is the most important to the network.

Degree

The traditional definition of the degree centrality is the number of edges that are connected to a node. Thus, the node with the highest number of edges connected to it, or the highest degree, is the most central to the network under this metric. We selected the degree centrality to analyze our network to show how connected a food desert is to the public transportation system. Thus, we identify how many bus stops are within a particular radius

of the representative address in each food desert, as bus stops allow access to the edges (public transit routes). To do this we used the 'Create Buffer' tool in ArcGIS Pro to define a radius of a certain distance around the representative addresses. Then using the 'Summarize Within' tool we determined how many bus stops fell in that radius. We considered 3 different distances: 0.25, 0.5, 0.75 miles based on [30].

Closeness

Traditionally, the closeness centrality examines the path distance from one node to all the other nodes in the network. Thus, the lower the closeness the more central a node is to the network. This definition of closeness is not particularly useful for these analyses, as we are not interested in how close food deserts are to one and other or how close grocery stores are to one and other; we are interested in identifying the distance between food deserts and grocery stores. Further, since we are exploring the role of public transportation, we are less interested in the distance between the two types of nodes as we are in the travel time between them (since public transit routes will rarely go directly from the starting point to the destination). Thus, our closeness centrality evaluates the number of stores that are within a certain total travel time from each food desert, where total travel time is defined as time spent walking and on public transit. Therefore, the higher the closeness, the more access a food desert has to grocery stores. This allows us to explore how connected food deserts currently are to grocery stores. We accomplished this using the Service Area layer, which is a part of the 'Network Analyst' tools. The Service Area layer shows the area reachable by public transit and walking within a given amount of time from a starting location on a particular day and time. Due to differences in the format of the GTFS data for the cities, the Service Area Layer was calculated for a general Wednesday at 7:00 pm for Lawrence, Albuquerque, Charlotte, and Raleigh and for the specific date of Wednesday, November 17th 2021 at 7:00 pm for Brown Deer. We ran the simulation with three different cutoff times for total time traveled: 10, 20, and 30 minutes. After the creation of the service area maps for each time cutoff, we used the Summarized Within tool to determine how many stores are reachable within each service area.

In addition to defining closeness to be a centrality measure for the food desert nodes, we also defined a related centrality measure for the grocery store nodes. This centrality measure, called the overlap centrality, looks at how many food desert service areas a grocery store resides in. Although not a traditional centrality measure, we consider this to be informative because the higher a store's overlap the more important it is to the network for food access. We defined an overlap of 1 to mean a store is only reachable by one food desert, an overlap of 2 to mean a store is reachable by two food deserts, and so on. We calculate this centrality measure using the data generated from calculating closeness. In addition to determining the number of stores reachable in a service area, the Summarize Within tool can also generate the list of stores. We were then able to see which stores appeared in more than one food desert's list. Since the overlap centrality is particularly interesting for grocery stores with an overlap of 2 or more, we omitted all stores with an overlap of 1 from the results shown in Appendix [A](#).

4.2.4 Sensitivity of Representative Addresses

While we selected representative addresses manually such that the address would be both in high population density areas and near the center of the food desert, there are many different methods for selecting representative addresses. Thus, we conducted a sensitivity analysis on how the representative addresses were chosen to determine the impact of these choices on the results of the degree and closeness analyses. To determine the sensitivity we conducted the analysis on Brown Deer and Lawrence by computing the degree and closeness centrality measures using two additional methods for selecting representative addresses. Then we employed a Wilcoxon Signed-Rank Test to determine if there was a statistically significant effect on the results depending on how the representative addresses were selected. The two additional methods we used to select representative addresses were: 1) randomly selecting a representative address from a list of high population density addresses (we will call this the 'random' method) 2) selecting an address in the center of the food deserts regardless of whether it is a residential area (we will call this the 'center' method). We will call the original method we used for selecting representative addresses the 'original' method. For the random method we used Reference USA/Data Axle Reference Solutions database to generate a list

of apartments for each city. This was done by searching for the following apartment related SIC codes: 651304, 651303, 651307, and 651398. Since not all food deserts have apartments in them, we also included addresses of public K-12 schools, as these tend to be located in or near residential communities. After eliminating addresses not located in food deserts, the remaining were assigned a number. Using a random number generator, one address was selected as the representative address for each food desert. If the randomly chosen address was an apartment complex the address was searched on Google to ensure it was accurate. If the address did not lead to an actual apartment, it was thrown out and a new address was randomly selected using the same method as before. For the center method, we dropped a pin in the center of the food desert (where street addresses exist) on ArcGIS Pro and those addresses were selected as the representative addresses. The addresses and results of the degree and closeness centrality analyses calculated with representative addresses from the random method and center method are provided in [Appendix A.1](#).

To then determine how sensitive our results are to the representative addresses, we used the Wilcoxon Signed-Rank Test. This is a non-parametric test for two populations when the observations are paired, and it conducts a paired difference test of repeated measurements on a single sample to assess whether their population mean ranks differ. We conducted this test pairwise between the results for the original method's representative addresses and the random method's representative addresses and the original method's representative addresses and the center method's representative addresses, as we are interested in how the results in this chapter could be affected. We ran the test for each distance/time cutoff. To do this analysis we used MATLAB's signrank function. We calculate the p , h , and stats values, where p is the p -value of a two-sided Wilcoxon Signed-Rank Test, h indicates a rejection of the null hypothesis or a failure to reject the null hypothesis at the 5% significance level, and stats returns the test statistic, W . Note that $h = 0$ represents failure to reject the null hypothesis and $h = 1$ represents rejection of the null hypothesis, where the null hypothesis is a zero median for the difference between paired samples.

4.3 Results

Figure 4.1 is the full network map for Brown Deer, WI produced by ArcGIS Pro. The full network maps for the other study sites can be found in Appendix A. The map depicts the 6 census tracts identified as food deserts in the Northwest portion of the city and the majority of bus stops and grocery stores located in the central and Southeast portion of the city.

4.3.1 Degree

Figure 4.2 not only shows the results of the degree analysis, but also the maps of all five study sites with the food deserts, representative addresses, bus stops, and the 0.75-mile buffer displayed. The full degree analysis results can be found in Tables 7-11 in Appendix A. The food desert with the highest degree within 0.75 miles across all the study sites is Raleigh's 700 with 73 bus stops and the second highest is Albuquerque's 900 which has 60 bus stops. In Lawrence, all representative addresses have access to at least one bus stop within 0.25 miles, which is the smallest distance we considered. Conversely, 50% of representative addresses in Brown Deer have a degree of 0 within 0.25 miles and the same is true for 47% of representative addresses in Albuquerque, 32% in Charlotte, and 44% in Raleigh. Across all the food deserts, Charlotte's 006 and 618 and Raleigh's 713 and 809 have the least access to their public transit systems as they have a degree of 0 within 0.75 miles. Thus, Raleigh has the largest disparity in public transit access for food deserts. On the other hand, Lawrence has the smallest disparity in public transit access as the largest degree within 0.75 miles is 22 and the smallest degree is 14. Across all the study sites, 11-20 is the most common number of bus stops within a 0.75 mile radius and this holds true on the city level for Brown Deer, Lawrence, and Charlotte. In Albuquerque 31+ is the most common number of bus stops within a 0.75 mile radius, while in Raleigh 1-10 bus stops is the most common. Table 4.2 and Table 4.3 show the results of the Wilcoxon Signed-Rank Tests between the original method's results and the random method, and the original method and the center method for Brown Deer and Lawrence. Note, since $h = 0$ for each test, we fail to reject the null hypothesis. Thus, there is zero median for the difference between paired samples.

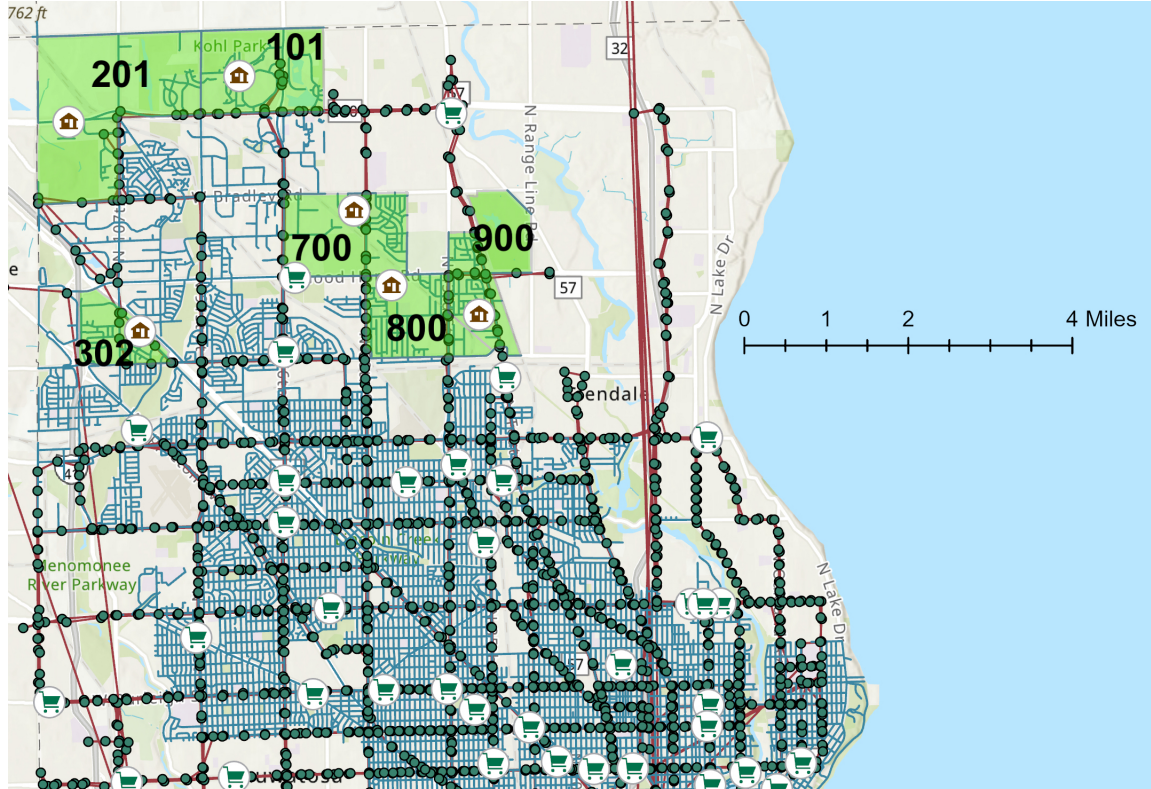


Figure 4.1: Full Network for Brown Deer. Network model for Brown Deer, WI Note: The food desert census tracts are shaded in green and labeled by the last three digits of their census tract number. The representative addresses are symbolized by a white circle with a picture of a house inside and the grocery stores are symbolized by a white circle with a shopping cart inside. For the edges, the bus stops are marked by small green circles and the streets are the blue lines. The maroon lines represent the transit lines but note they do not always follow the streets or look like a typical map of the transit system. This is because they are representative of logical connections in the transit system, which is used by The 'Public Transit Evaluator' tool to determine travel times [31].

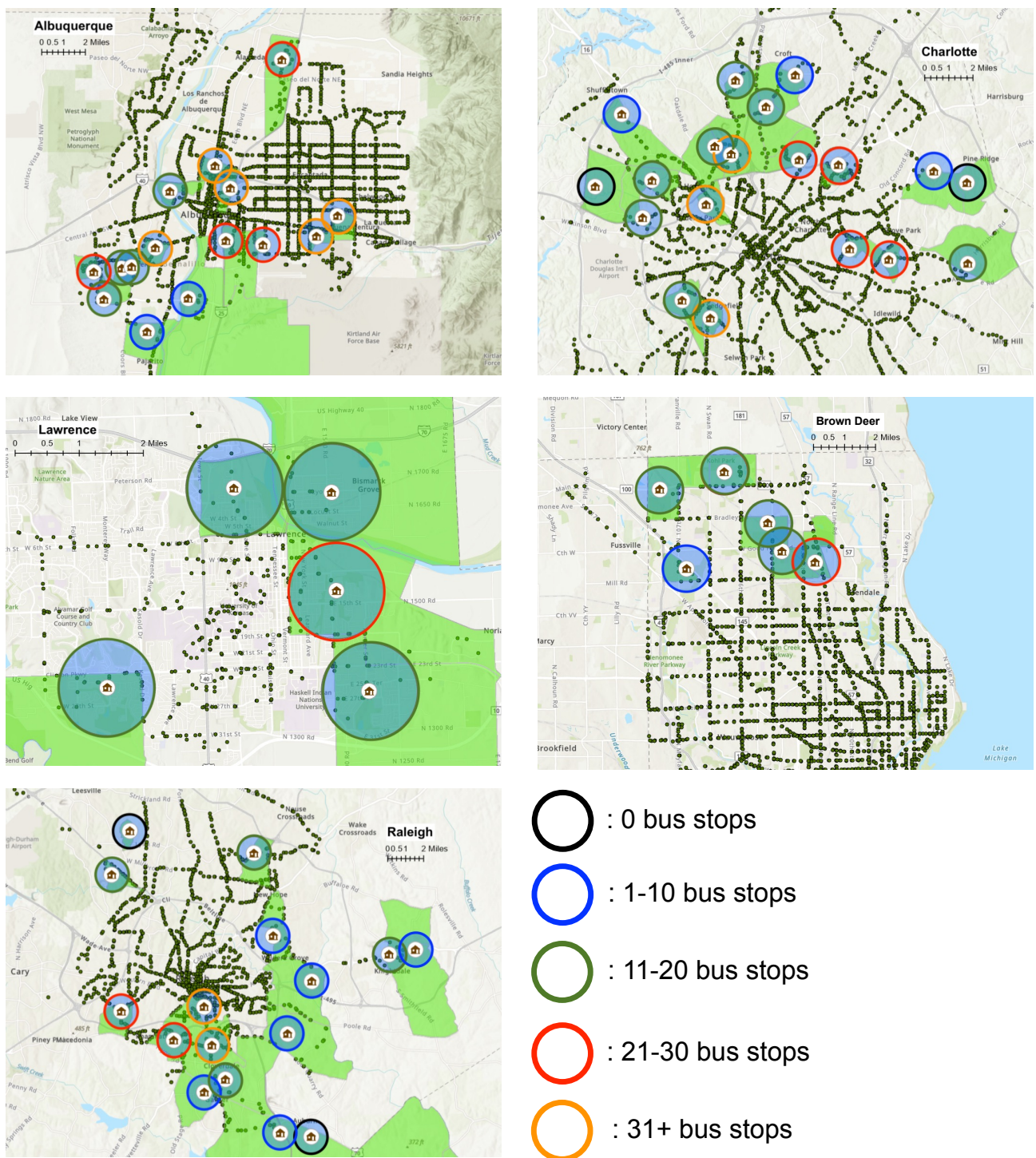


Figure 4.2: Degree Results. Maps of all five study locations with food desert shaded in green, the representative apartments marked by a white circle with a house inside, bus stops represented by small dark green circles, and the 0.75 miles radius circle shaded in light blue outlined with the color associated to the number of bus stops that fall inside it.

Table 4.2: Wilcoxon Signed-Rank Tests for Degree in Brown Deer. Results of the Wilcoxon Signed-Rank Tests between the original method and the random method and between the original method and the center method's degree analysis results in Brown Deer.

	Original vs. Random	Original vs. Center	Original vs. Random	Original vs. Center	Original vs. Random	Original vs. Center
Distance (in minutes)	<i>p</i>	<i>p</i>	<i>h</i>	<i>h</i>	Stats	Stats
0.25	0.5625	0.8125	0	0	5	6
0.5	0.0625	1	0	0	0	11
0.75	0.2812	0.4375	0	0	5	6

Table 4.3: Wilcoxon Signed-Rank Tests for Degree in Lawrence. Results of the Wilcoxon Signed-Rank Tests between the original method and the random method and between the original method and the center method's degree analysis results in Lawrence.

	Original vs. Random	Original vs. Center	Original vs. Random	Original vs. Center	Original vs. Random	Original vs. Center
Distance (in minutes)	<i>p</i>	<i>p</i>	<i>h</i>	<i>h</i>	Stats	Stats
0.25	0.2500	0.2500	0	0	3	13
0.5	0.1250	0.1250	0	0	1	10
0.75	0.1250	0.1875	0	0	1	13

4.3.2 Closeness

Figure 4.3 shows the closeness results for all 61 food deserts across the different study sites. Tables 12-16 in Appendix A provide additional details for the information in Figure 4.3 for each study site. Across all five cities, Albuquerque’s 707 and 901 food deserts have the highest closeness with 22 and 9 grocery stores within 30 minutes, respectively. The food deserts with the highest closeness within 30 minutes for each city is as follows: Brown Deer’s 900 with 7 grocery stores, Lawrence’s 004 with 4 grocery stores, Charlotte’s 603 with 5 grocery stores, and Raleigh’s 707 with 6. The lowest closeness a food desert can have is zero grocery stores within a total travel time of 30 minutes. Out of all the representative addresses across the 5 study sites, 20 (or about 33%) have a closeness of zero. Raleigh has the most, and highest proportion of, food deserts without grocery store access within 30 minutes at 44%, while Albuquerque and Lawrence have the lowest proportion with 20% each. Brown Deer and Charlotte fall in between with 33% and 37%, respectively. Four representative addresses have access to at least one store within 10 minutes (900 in Brown Deer, 201 in Charlotte, and 101, 301 in Raleigh) and out of the food deserts that have grocery store access, just over half must travel at least 20-30 minutes to access a grocery store. In addition, three of the five study sites have grocery stores with overlaps greater than 1, meaning the store is reachable by more than one food desert within 30 minutes: Brown Deer, Albuquerque, and Raleigh. Albuquerque has the most with 10 grocery stores with an overlap greater than 1, while Brown Deer and Raleigh both have 2. Of the 14 stores all but 3 have an overlap of 2 food deserts; 3 stores in Albuquerque and 1 in Raleigh have an overlap of 3 food deserts. Table 4.4 and Table 4.5 show the results of the Wilcoxon Signed-Rank Tests between the original method’s results and the random method and the original method and the center method for Brown Deer and Lawrence. Note, since $h = 0$ for each test, we fail to reject the null hypothesis. Thus, there is zero median for the difference between paired samples.

Five of the grocery stores across all of the study sites that appeared in the closeness analysis were flagged because we were unable to verify whether or not they provide fresh produce and meat (as indicated in the Methods section). One of the flagged stores was reachable by

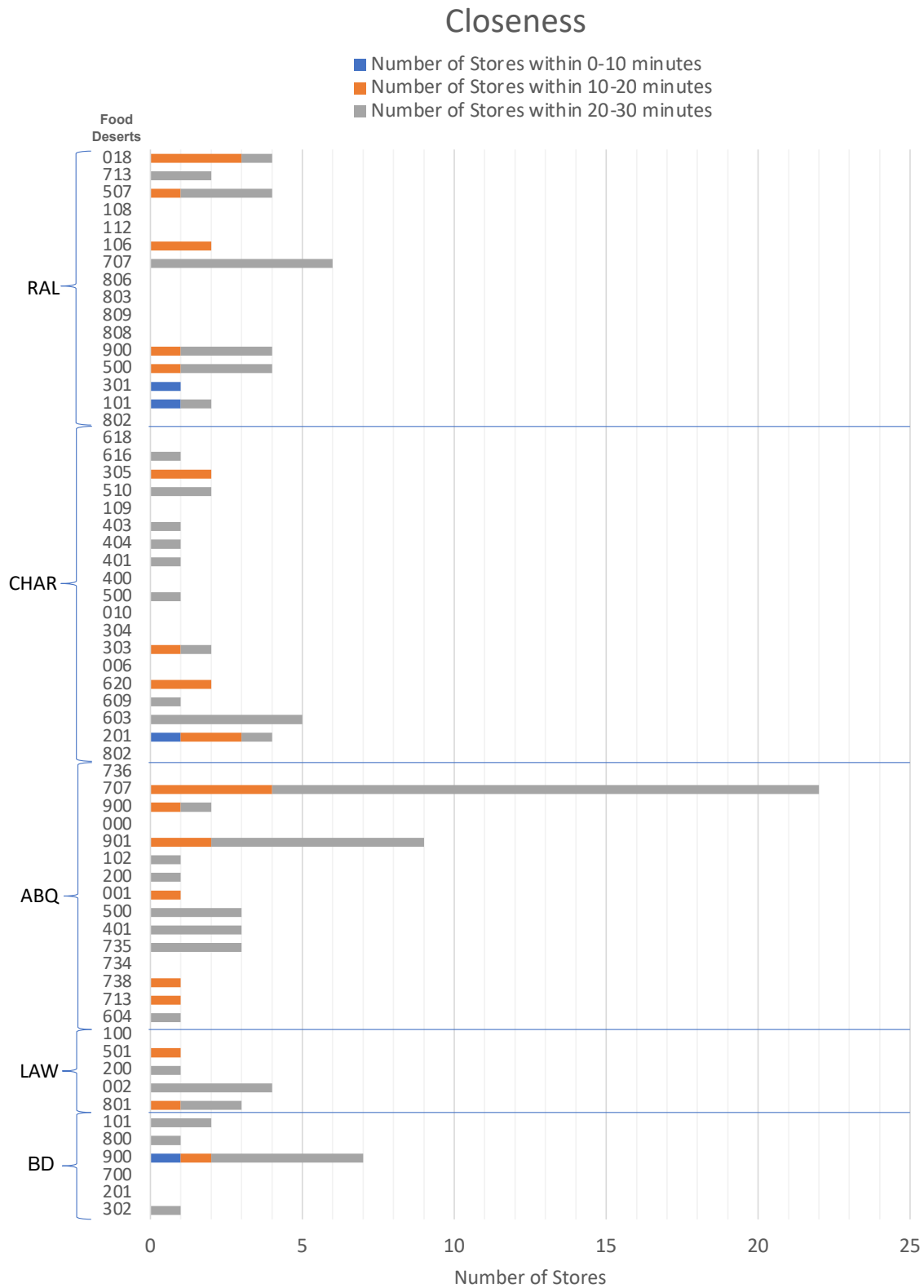


Figure 4.3: Closeness Results. The number of grocery stores within 0-10, 10-20, 20-30 minutes of total travel time for all of the representative addresses across the study sites. BD-Brown Deer; LAW-Lawrence; ABQ- Albuquerque; CHAR-Charlotte; RAL-Raleigh

Table 4.4: Wilcoxon Signed-Rank Tests for Closeness in Brown Deer. Results of the Wilcoxon Signed-Rank Tests between the original method and the random method and between the original method and the center method's closeness analysis results in Brown Deer.

	Original vs. Random	Original vs. Center	Original vs. Random	Original vs. Center	Original vs. Random	Original vs. Center
Distance (in minutes)	p	p	h	h	Stats	Stats
0-10	1	1	0	0	1	1
10-20	1	1	0	0	1	1
20-30	0.2500	0.2500	0	0	6	13

Table 4.5: Wilcoxon Signed-Rank Tests for Closeness in Lawrence. Results of the Wilcoxon Signed-Rank Tests between the original method and the random method and between the original method and the center method's closeness analysis results in Lawrence.

	Original vs. Random	Original vs. Center	Original vs. Random	Original vs. Center	Original vs. Random	Original vs. Center
Distance (in minutes)	p	p	h	h	Stats	Stats
0-10	1	1	0	0	0	0
10-20	1	1	0	0	1	1
20-30	0.7500	0.3750	0	0	4.5	8.5

the 900 food desert in Brown Deer and one was reachable by the 901 food desert and another one by the 707 food desert in Albuquerque. In Charlotte the 305 food desert could reach a flagged store, and the same was true for the 018 food desert in Raleigh. Due to the low number of flagged stores appearing in each city’s closeness analysis, the results would not be substantially different if the stores had been removed from the list. While including the flagged stores in the analysis had a relatively small impact, performing the closeness analysis on the unverified grocery store list has a larger impact. Across all the food deserts in each city, the sum of the closeness results for the number of grocery stores reachable within 30 minutes is increased by 73 when computing the closeness using the unverified list compared to the verified one. This breaks down to an increase in the closeness results across all the food deserts in Brown Deer by 12, in Lawrence by 5, in Albuquerque by 27, in Charlotte by 11, and in Raleigh by 18. The full results of the closeness analysis done with the unverified list is provided in Appendix [A](#).

4.4 Discussion

We used the degree centrality measure to quantify the number of bus stops within different buffer distances (0.25, 0.5, 0.75 miles) of representative addresses in each food desert. While a high degree centrality does not guarantee reasonable access to fresh and nutritious food, it does give an idea which food desert nodes are most connected or central to the network and thus more likely benefit from access to public transportation. This access to public transit could potentially lead to access to grocery stores that are further away. City leaders can use this centrality measure to decide where to add more bus stops or alter/create bus lines that will help connect low-degree census tracts to public transportation and improve access to food. This is especially important for food deserts that have no, or very low, access to public transit. Further consideration should be taken to ensure bus stops near the food deserts provide direct access to a grocery store (e.g., require no or minimal bus transfers) and that there are an adequate number of bus stops near grocery stores/business districts. Access to good public transit is important, especially in low-income neighborhoods, where other forms of reliable transportation may not be available. Lack of access to public transit

can be especially concerning in census tracts that have low vehicle access (tracts in which more than 100 households have no access to a vehicle and are more than 1/2 mile from the nearest supermarket [88]), such as all of the food deserts in Brown Deer, Lawrence's 200, Albuquerque's 000, 200, 102, 901, and 707, Charlotte's 304, 305, 400, 401, 403, 404, 500, 603, 609, 610, 802, and Raleigh's 018, 101, 500, 803, and 900.

Determining what a high degree or adequate access to public transit means will vary across cities and depend on the city and their goals. For instance, in Lawrence, every food desert has a bus stop within 0.25 miles, but the most common number of bus stops within 0.75 miles is 11-20, while in Albuquerque the most common number of bus stops within 0.75 miles is 31+, even though 47% of Albuquerque's food deserts have no bus stops within 0.25 miles. An argument can be made for both locations for which city has a higher degree, and while the best option is to have the most bus stop access at the closest distance, that may not be immediately feasible. Thus, taking various factors into consideration such as existing infrastructure, budget, and population density, city leaders can decide on how best to increase the degree of their food deserts and thereby increase potential access to food.

While the degree centrality focuses on how connected a food desert is to the public transit system, closeness examines how connected a food desert is to grocery stores through the public transit system. We chose to run the closeness analysis for Wednesday at 7:00 pm because this is after the traditional workday ends but is late enough that it will avoid major rush hour traffic that typically increase travel times. Note, that these cutoff times represent total time traveled, which includes walking plus public transit. This is important to consider because walking long distances presents different challenges than riding a bus, such as harsh weather conditions, safety, accessibility, and carrying heavy groceries. Analyzing closeness is valuable for mitigating harm caused by food deserts because it shows which food deserts need to be focused on first. While ideally cities would be able to increase food access for everyone in food deserts, those that have a low closeness, and especially those without any access to grocery stores within 30 minutes, need to be addressed first. This could look like enacting short-term solutions in those areas immediately, such as mobile food pantries or

providing rides to grocery stores, while more permanent solutions are implemented. Further, by using a network model to address these issues, simulations can be run to help determine where routes could be modified or added to connect food deserts to previously unreachable grocery stores or make access to them quicker. Studying this via modeling is much more efficient for cities than making changes to their transit systems and studying the realized effects only after implementation. While ideally closeness should be used to increase food access, at the least it should be used to ensure access won't be made worse by changes to the public transit system. Thus, a closeness analysis should also be run when any adjustments are made to the public transit system, even if the changes aren't directly related to food access.

While city leaders don't have complete control over where grocery stores open, they can influence and incentivize businesses to open in certain locations. Thus, considering an overlap centrality can help maximize the impact city leaders have on grocery stores. Note that currently not many stores have an overlap greater than 1, so most grocery stores are not in ideal locations for increasing food access. Thus, city leaders should try to incentivize stores to open in locations where they would have a higher overlap, so food access can be increased for multiple food deserts at once. Note, that spacing out grocery stores such that they have a low overlap may be a purposeful strategy by the grocery stores to maximize profit, so urban planners and economist may need to investigate this further. By knowing the overlap of current or proposed stores, it could help city leaders make decision about allocation and distribution of resources. For example, stores with a higher overlap could be more likely to receive government aid or be of higher priority for the necessary licensing and other bureaucratic procedures. Further, if any grocery stores with a higher overlap close, city leaders will be aware that the location is an important one for food access and could try to incentivize a new grocery store to open in a similar location.

When considering the relationship between degree and closeness, we would not necessarily expect to see any correlation between the two as a mathematical property of an arbitrary network. Although, when considering a relationship between them as a property of networks

that describe human communities, it may be intuitive to think the higher a food desert's degree is the higher their closeness should be (since they have more access to public transit and therefore better access to grocery stores that are farther away); that is not necessarily the case (although, Brown Deer and Lawrence do seem to follow this trend). On the other hand, in general, we do see if a food desert has a lower degree, they tend to also have a lower closeness. For example, of the 4 food deserts that have a degree of 0 within 0.75 miles, 3 don't have access to grocery stores within 30 minutes. One has access to two grocery stores within 20-30 minutes mostly by walking (since they don't have access to a bus stop within 0.75 miles). However, this is not a bi-directional relationship. Thus, low degree usually implies lower closeness, but low closeness does not imply low degree. So, based on our study sites we can conclude that degree does not necessarily have a positive effect on grocery store access. This may imply cities are currently under utilizing their public transit systems as a way to increase food access. Many food deserts have decent access to public transit, but the public transit is not connecting them to grocery stores. For future public transit planning, officials should explicitly take food access into account. This could be as simple as adding bus stops near grocery stores on existing routes or more complicated such as adding new or modifying existing lines such that they reach grocery stores. To get a better understanding of the relationship between degree and closeness more city-specific factors may need to be considered and evaluated individually.

For the sensitivity analysis of the representative addresses for both Brown Deer and Lawrence, all the Wilcoxon Signed-Rank Tests show that there is no evidence that the choice of representative addresses, or the method used to select representative addresses, has a statistically significant impact on the overall structure of the results. Thus, while the individual numerical results of the degree and closeness analysis may change depending on the method used for selecting representative addresses, the overall structure of the results (e.g. highest degree/closeness to lowest degree/closeness) is unlikely to be impacted. (Note that we only performed this sensitivity analysis on Brown Deer and Lawrence and determined that, since there was no evidence of impact, these initial investigations could safely forgo

running the same analysis on the other study sites at this stage; this presents an opportunity for future exploration.)

4.4.1 Limitations

The primary limitation of this study was that the database we used for the grocery store data is inaccurate. Thus, moving forward, having access to a more complete data set would improve the accuracy of the models and results. As mentioned above, to correct the inaccurate grocery store data we performed a verification process that was mostly done by hand. During this process, finding adequate evidence that a business was a grocery store was challenging. Also, since this was done manually, human error and bias is always a concern. Further, while we know that the original grocery store lists contained businesses that did not satisfy our definition of a grocery store, we were not able to determine if businesses that would satisfy the definition were left out, or misclassified, by the database. Another limitation to our analysis is working with representative addresses. Although we showed that different methods for selecting representative addresses did not cause a statistically significant change in the average ranks of the food deserts results for the closeness and degree analysis within each study site, it can affect the numerical values of the results. Thus, if city leaders want to employ this model in their cities, it will be important to determine the representative addresses in a manner that is most useful for their efforts and desired outcomes.

As mentioned previously, there are many different types of centrality measures that can be used to evaluate a network. While we purposely selected the degree and closeness centrality measures for the information they provide about the connectivity of food deserts to both the public transit system and the grocery stores, along with painting a realistic picture for the current state of food access in food deserts, our results and recommendations may have been different if different centrality measures had been implemented. Our results are a direct consequence of the centrality measures we chose. Further, our selection of centrality measures is limited by the capabilities of ArcGIS Pro; specifically we could not implement analysis of betweenness centrality. Betweenness is a centrality measure that evaluates the importance of efficient paths. We wanted to explore the betweenness centrality in our networks to determine

which bus routes are most heavily relied on to connect food deserts to grocery stores. We planned to calculate the shortest route between each food desert to all the grocery stores, then analyze which bus route was used the most often for these connections. This could have provided important information to city leaders about which bus routes are vital to food access and influence how those routes are modified in the future. Unfortunately, we were not able to determine a way to calculate betweenness on ArcGIS Pro. The software currently lacks the capabilities of providing public transit routing directions, as provided by companies like Google Maps. Finally, this study focuses on access to the closest grocery stores, but this may not be realistic as people have grocery store preferences and will often travel further to reach a desired grocery store. So, in the future incorporating different grocery store characteristics, like pricing, product range/availability, and location safety into the model could offer better insight into the real distance people travel for groceries.

4.5 Conclusion

Analyzing a complex situation like food deserts with a network model and centrality measures can be an incredibly useful tool for city leaders. It allows for adjustments to be made in the model and simulations to be run, before they are implemented in the public transit systems, which is a time- and cost-effective way to consider issues surrounding food access and deserts. This lets city leaders tackle these issues dynamically. As changes are made, analysis can be run to offer adjustments before the next set of changes are implemented. This is especially important as individual behavior and grocery stores are likely to react in real time to changes. In this chapter, we used network models to characterize the role public transportation has in connecting food deserts to grocery stores in 5 U.S. cities. Our network models focused on two centrality measures: degree and closeness. Results of the degree centrality measure varied across study sites; 1 study site (Lawrence, KS) had at least 1 bus stop within 0.25 miles of the representative address for each food desert. Conversely, 2 study sites (Charlotte, NC and Raleigh, NC) each had 2 representative addresses with 0 bus stops within 0.75 miles. When using the closeness centrality measure, 2 food deserts in Albuquerque, NM had the highest number of grocery stores within 30 min (22 and 9), while

44% of food deserts in Raleigh, NC had 0 grocery stores within 30 minutes. These results will be useful to city leaders and planners in developing strategies to use public transportation to improve food access to residents of food deserts in their cities.

While this study offers suggestions for city leaders for how to best utilize their public transit system to increase food access, the next step is to implement these suggestions in a city's model to see the impact it would have on the centrality measures. A theoretical example of analyzing a food desert network when various changes are made to the transit system is explored in Appendix [A.2](#). Further, we could offer more impactful suggestions for how to modify public transportation by conducting a sensitivity analysis between food deserts that have low vehicle access and those that do not. This would offer insight into where public transit plays a more important role in food access, and which areas need to be prioritized for transit improvements, as people who have reliable vehicle access are less likely to rely on public transportation to reach grocery stores. More generally, this work could help develop a new definition of food deserts that takes public transportation into account. Right now, even if a food desert has adequate access to grocery stores through public transit, it is still considered a food desert by the original definition. Thus, by having a definition that takes public transit into consideration, cities will be able to more easily identify census tracts that are most isolated from food access and focus their resources and support where it is needed most. In addition to studying access to grocery stores, network models like this can be employed by city leaders to study the role public transportation plays in accessing other essential services and resources, such as employment, medical care, and education. While network modeling has often been applied to study public transportation and other model types (such as agent-based models) have been used to study individuals' behaviors around public transit use and food access, this chapter combines these two ideas and uses network modeling to explore food access through public transit. This novel approach to understanding the role public transportation plays in food access allows for suggestions to be made to city leaders on how to best utilize their public transit system to help alleviate the burden of food deserts.

Chapter 5

Conclusion

The overall goal of this dissertation is to use mathematical modeling to understand and help mitigate societal issues that communities face. To accomplish this goal we focus on building up the current toolkit available in applied mathematics, specifically by expanding the use of network theory, to tackle social and community problems. By considering three community-level issues, disease spread, violence, and food access, we are able to explicitly show the utility of applying mathematical modeling to social issues. In addition to contributing to the body of mathematical knowledge, we aimed for the results to be practically useful for those in other fields, especially community leaders and decision makers. In this chapter we will review the major findings from each project and see how the results answered the research questions posed in Chapter 1. Then we will step back and discuss how Chapters 2, 3, and 4 help achieve the overall goals of the dissertation. Next, we will consider some of the limitations and challenges that we faced. Finally, we offer some future directions to continue the research we have started and end by highlighting the some main outcomes from this dissertation.

The work in Chapter 2 is motivated by the question: is it possible to develop a new algorithm that is both equivalent to the current state of the art method, the Next Generation Matrix method, and more accessible to a broad audience? The simple answer to our question is: yes, we were able to develop an equivalent method using flow networks, called the Max-Flow Min Cut method. By relying on flow networks instead of differential equations and linear

algebra, it is more accessible to a non-mathematical audience as networks are more intuitive. We not only showed the Max-Flow Min-Cut method is equivalent to the Next Generation Matrix method, but also provided detailed step-by-step instructions for its implementation. Further, we presented multiple general examples and specific examples from the literature using our new method. In addition to being more accessible, the Max-Flow Min-Cut Method also provides novel insights about the basic reproductive number and prevention strategies by analyzing the cut-sets. It can also be applied to systems that have a non-unique endemic equilibria and provide non-equilibrium behavior estimates by using continuous time dynamics and continuous time with delays in the flow network.

The research question for Chapter 3 is: by building upon the emerging framework of treating violence as an infectious disease and viewing it as a public health issue, can we use traditional epidemic models to learn about the dynamics of violence spread and come up with viable solutions for how to reduce the effects of it in our communities? We were able to answer our question by first expanding the analogy between violence and infectious disease. This included incorporating a within-host component to the analogy. We then developed both a One Exposed Class and a Three Exposed Class susceptible-exposed-infected-recovered epidemic model to study violence spread through exposure in a community. Performing equilibrium and stability analysis we found for the One Exposed Class Model that the endemic equilibrium is stable and that the Three Exposed Class Model is too complicated for traditional stability and equilibrium analysis. Thus we explore its dynamics using numerical simulations. We examine six scenarios: a base case, increased propensity for violence after exposure, increased intervention measures, increased support after exposure, a decrease in intervention measure after they have been in effect, and delayed introduction of increased intervention measures. These simulations showed that the negative effects of violence exposure appear to be more impactful than the positive effects of preventative measures in a community and that, therefore, investing in violence intervention measures is much more effective than investing in support after exposure.

Based on these results we offered some suggestions for how to reduce the effects of violence in our communities. First we noted that investing in protective factors (which offer support to individuals so they can better cope with a violence exposure) are more effective at preventing the spread of violence than interruption/disruption techniques (which help decrease a person's propensity to commit violence after it as been increased due to exposure). Also, that preventative measures need to be introduced to a community early and consistently to maximize their effects. Further, we defined a basic reproductive number-type quantity for violence and other epidemiologically-based social models called the contributive number. By formulating the system into a flow network, we can calculate the flow of damage (due to violence) through a community using the method developed in Chapter 2. Thus, the contributive number can help capture the amount of damage an act of violence can do to a community.

Chapter 4's research question is: can public transportation networks be used to help alleviate food deserts? We discovered the short answer to this question is yes, but currently the cities we studied do not seem to be using their public transit systems effectively to that end. By comparing the degree and closeness results, we note that high degree does not generally imply high closeness. This means even when food deserts are well connected to their public transit system, it doesn't always guarantee access to grocery stores. This implies there may be more effective ways to utilize public transit to connect food deserts to grocery stores. To help use a city's public transit systems more effectively we offered policy suggestions based on our analysis.

The degree analysis can help city leaders decide where to add more bus stops, or alter and create bus lines, by identifying low-degree food deserts. In addition to adding bus stops near food deserts, degree should also be considered near grocery stores and business districts to ensure increased public transit access does increase food access. By considering the closeness analysis, city leaders can prioritize low-closeness food deserts, especially those without any access to grocery stores within 30 minutes, in receiving both short term aid and long term solutions. Closeness simulations can also help determine where routes could be altered to

connect food deserts to new grocery stores or make access to stores quicker. The overlap centrality can help decision makers know where to incentivize grocery stores to open that would increase food access for the most food deserts at once. Taking these policy suggestions into account will help city planners better utilize public transit to alleviate the burden of food deserts.

As we have seen, in all three projects we have been able to successfully apply mathematical modeling to study community level issues, which is our main goal. Further, while these models have yet to be applied in real communities, we are confident based on the results provided, that the use of mathematical modeling can aid in alleviating the community issues explored here. In addition to all three problems being mathematically interesting, the discussion and results are directly applicable to those working on the problems on the ground. In chapter 2, we provide a user-friendly guide for how to apply our new method to calculate and interpret the basic reproductive number. In chapter 3 we show the benefits and types of results that can be expected by employing mathematical models in support of violence prevention and, in Chapter 4, we offer detailed models and suggestions on how to better utilize the public transit system for increasing food access. Overall, we are able to offer concrete policy suggestions to community leaders and decision makers based on our analysis. This is especially true in chapters 3 and 4. While Chapter 2, doesn't offer specific policy suggestions itself, it does provide a new method for public health officials (and their in-house analysts) to utilize when making their policy suggestions. Thus, we accomplish our goal of having results that are also useful to those in fields other than mathematics.

Further we have been able to build upon and expand the use of network theoretic methods as tools to address these societal-based community issues. Network theory is an important aspect of all three chapters and each chapter utilizes network theory in a unique way. Chapter 2 employs a flow network for the calculation and analysis of the basic reproductive number and Chapter 3 then builds on this by applying this flow network-based method to study a basic reproductive number-type concept for violence. Chapter 4 takes a more traditional approach to modeling with networks and analyzing centrality measures and applies it to the

novel application of food access through public transit. Therefore, not only has each chapter answered its research question, but the dissertation as a whole is able to accomplish the goals we set.

Even though we are able to accomplish our goals, this work is not without its limitations and challenges. In general, mathematical models are limited as they must balance capturing the complexity of the real-life system while being simple enough that useful analysis can be done. By analyzing and modeling social and community issues, we are inherently incorporating humans and human behavior into our models, which is notoriously challenging. Disease spread, violence, and public transit use are all dependent on human behaviors and decisions. Thus, every project presented here made simplifying assumptions. While we believe that the assumptions made were justified and satisfy this balance, they do introduce limitations to the realism of the models and should be interpreted acknowledging those implications. Another limitation is that the models in Chapters 3 and 4 require data, either now (Chapter 4) or in the future to calibrate the model (Chapter 3), which, as seen in Chapter 4, can present challenges. In order for these models to be of greatest use, the data needs to be accurate, which may be difficult to find. Due to the sensitive nature of Chapter 3, data for this model may be especially difficult to access and consideration needs to be taken for any implicit biases that may be embedded in the data. Moving forward, an important aspect that could limit the application of the results is the acceptance of mathematical modeling as a useful tool by experts in other fields. While mathematical techniques and concepts may already be widely used and accepted in some fields, like epidemiology, in others, like violence prevention and food security, tools like mathematical modeling may not be as common and thus receive more resistance. In order for the ideas and projects presented here to progress it will be important to promote the acceptance of mathematical modeling and clearly explain its utility as a new tool for addressing societal problems.

The projects presented in this dissertation offer many opportunities for continued development, expansion, and future directions. In general, an important next step will be to work with community leaders in applying the models and techniques we developed here to

their communities. By applying the methods in the real world, we will not only be able to see the true scope of the tools applicability, but also make any adjustments needed to increase their utility. More specifically, the Max-Flow Min-Cut method from Chapter 2 can be expanded by increasing the types of models it can be applied to and for Chapter 3 two important next steps are to incorporate data and integrate the epidemic model into an Immuno-Epidemiological framework, including formulating the within-host model. In Chapter 4, collaborating with policy makers from a city that wants to address food access through their public transit would allow us to apply our suggestions to the city's model and test how food access improves. While we develop the framework and tools here to investigate specific issues, they could also be applied to related problems. The foundation of Chapter 3, applying epidemic modeling to violence, can be applied to related social issues, such as the spread of conspiracy theories, mis- or disinformation, and political ideologies, to name a few. The network framework used in Chapter 4 could easily be adjusted to study access to employment or other social services instead of grocery stores. Overall, we hope this dissertation will help inspire future applications of modeling to social and community issues.

While this dissertation is by no means the first instance of applying mathematics and mathematical modeling to study social and community based problems, it does offer many novel and promising insights. We are able to apply traditional modeling tools, like dynamical systems and network theory, to gain a new perspective on social issues. By looking at the well-studied topic of the basic reproductive number from a network theoretic perspective in Chapter 2, we were able to develop a new and more accessible method for its calculation and interpretation. In Chapter 3 we applied a variation of the classic susceptible-exposed-infected-recovered epidemic model to study the spread of violence through exposure in a community, which is a novel approach. Chapter 4 employs a standard network/graph model to explore the role that public transportation systems play in connecting food desert residents to grocery stores. While network models have been used to study public transit before, they had yet to be used to study the role it plays in food access. The motivation to look at community issues through the lens of mathematical modeling follows in line with the

motivation of the first mathematical model of smallpox and all of the models before that:
to help us gain a better understanding of, and hopefully improve, the world around us.

Bibliography

- [1] (2019). foodaccess2019 (mapserver). <https://gis.ers.usda.gov/arcgis/rest/services/foodaccess2019/MapServer>. 79, 147
- [2] (2021). Ku-lts-gtfs. <https://lawrencetransit.org/developer/>. 79
- [3] Abel, K. C. and Faust, K. M. (2020). Modeling complex human systems: An adaptable framework of urban food deserts. *Sustainable Cities and Society*, 52:101795. 77
- [4] Albuquerque, C. o. (2021). Streets. <https://www.cabq.gov/gis/geographic-information-systems-data>. 79
- [5] Althaus, C. L. (2014). Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa . *PLOS Currents*, 6. 9
- [6] Aslan, I. H., Demir, M., Wise, M. M., and Lenhart, S. (2020). Modeling covid-19: Forecasting and analyzing the dynamics of the outbreak in hubei and turkey. *medRxiv*. 37
- [7] Baláz, V., Kvasnička, V., and Pospíchal, J. (1992). Two metrics in a graph theory modeling of organic chemistry. *Discrete applied mathematics*, 35(1):1–19. 2
- [8] Becker, M. G., Hall, J. S., Ursic, C. M., Jain, S., and Calhoun, D. (2004). Caught in the crossfire: the effects of a peer-based intervention program for violently injured youth. *Journal of Adolescent Health*, 34(3):177–183. 61, 64
- [9] Benkö, G., Flamm, C., and Stadler, P. F. (2003). A graph-based toy model of chemistry. *Journal of Chemical Information and Computer Sciences*, 43(4):1085–1093. 2
- [10] Berestycki, H., Rodriguez, N., and Ryzhik, L. (2013). Traveling wave solutions in a reaction-diffusion model for criminal activity. *Society for Industrial and Applied Mathematics*, (Vol. 11, No. 4, pp. 1097–1126). 38
- [11] Bernoulli, D. and Blower, S. (2004). An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. *Reviews in medical virology*, 14(5):275. 1

- [12] Bonnasse-Gahot, L., Berestycki, H., Depuiset, M.-A., Gordon, M. B., Roché, S., Rodriguez, N., and Nadal, J.-P. (2018). Epidemiological modelling of the 2005 french riots: a spreading wave and the role of contagion. *Scientific Reports*, (8:107). 38
- [13] Brauer, F. (2009). Mathematical epidemiology is not an oxymoron. *BMC Public Health*, 9(1):1–11. 1
- [14] Brauer, F., van den Driessche, P., and Wu, J. (2008). *Mathematical Epidemiology*, volume 1945 of *Mathematical Biosciences Subseries*, chapter Further Notes on the Basic Reproduction Number, pages 159–178. Springer. 3, 9, 12, 13, 14, 16, 20
- [15] Castillo-Garsow, C. W. and Castillo-Chavez, C. (2020). A tour of the basic reproductive number and the next generation of researchers. *An Introduction to Undergraduate Research in Computational and Mathematical Biology*, page 87. 10
- [16] Cerdá, M., Tracy, M., and Keyes, K. M. (2018). Reducing urban violence: a contrast of public health and criminal justice approaches. *Epidemiology (Cambridge, Mass.)*, 29(1):142. 38
- [17] Coleman-Jensen, A., Rabbitt, M. P., Gregory, C. A., and Singh, A. (2021). Household food security in the united states in 2020. Report, Economic Research Service. 76
- [18] County, W. (2021). Streets in wake county, nc. <https://data-ral.opendata.arcgis.com/search?tags=Transportation>. 79
- [19] Daher, M. (2003). World report on violence and health. *Le Journal Médical Libanais. The Lebanese Medical Journal*, 51(2):59–63. 4
- [20] Dantzig, G. and Fulkerson, D. R. (2003). On the max flow min cut theorem of networks. *Linear inequalities and related systems*, 38:225–231. 12
- [21] Data, O. M. (2021). Open mobility data. <https://transitfeeds.com>. 79
- [22] David-Ferdon, C., Vivolo-Kantor, A. M., Dahlberg, L. L., Marshall, K. J., Rainford, N., and Hall, J. E. (2016). A comprehensive technical package for the prevention of

- youth violence and associated risk behaviors. Technical report, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, Atlanta, GA. 36
- [23] de Camino-Beck, T., Lewis, M. A., and van den Driessche, P. (2009). A graph-theoretic method for the basic reproduction number in continuous time epidemiological models. *Journal of Mathematical Biology*, 59. 10
- [24] Delgado, S. A., Alsabahi, L., Wolff, K. T., Alexander, N. M., Cobar, P. A., and Butts, J. A. (2017). The effects of cure violence in the south bronx and east new york, brooklyn. 61, 63, 64
- [25] Derrible, S. and Kennedy, C. (2009). Network analysis of world subway systems using updated graph theory. *Transportation Research Record*, 2112(1):17–25. 77
- [26] Development, C. o. M. D. o. C. (2019). Milwaukee fresh food access report. 77
- [27] Diekmann, O., Heesterbeek, J., and Metz, J. (1990). On the definition and the computation of the basic reproductive ratio, $\mathcal{R}_0$ in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, (28). 3, 4, 9, 10
- [28] Dietz, K. (1975). Transmission and control of arbovirus diseases. *Epidemiology*, 104:104–121. 3, 9
- [29] Eames, K. T. and Keeling, M. J. (2004). Monogamous networks and the spread of sexually transmitted diseases. *Mathematical Biosciences*, 189. 2, 10
- [30] El-Geneidy, A., Grimsrud, M., Wasfi, R., Tétreault, P., and Surprenant-Legault, J. (2014). New evidence on walking distances to transit stops: Identifying redundancies and gaps using variable service areas. *Transportation*, 41(1):193–210. 76, 83
- [31] ESRI (2021). Create and use a network dataset with public transit data. <https://pro.arcgis.com/en/pro-app/latest/help/analysis/networks/create-and-use-a-network-dataset-with-public-transit-data.htm#>. 87, 127, 128, 129, 130

- [32] Essam, J. W. (1971). Graph theory and statistical physics. *Discrete Mathematics*, 1(1):83–112. [2](#)
- [33] Felson, R. B. (2009). Violence, crime, and violent crime. *International Journal of Conflict and Violence (IJCIV)*, 3(1):23–39. [37](#), [66](#)
- [34] Finkelhor, D., Turner, H., Hamby, S., and Ormrod, R. (2011). Polyvictimization: Children’s exposure to multiple types of violence, crime, and abuse. *Juvenile Justice Bulletin*, (NCJ 235504). [36](#), [64](#), [66](#)
- [35] Firth, J. A., Hellewel, J., Klepac, P., Kissler, S., Group, C. C.-. W., Kucharski, A. J., and Spurgin, L. G. (2020). Using a real-world network to model localized COVID-19 control strategies. *Nature Medicine*, 26. [2](#), [10](#)
- [36] for Disease Control, C. and Prevention (2019). Preventing adverse childhood experiences: Leveraging the best available evidence. Technical report, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, Atlanta, GA. [40](#), [64](#), [65](#)
- [37] Ford, L. R. and Fulkerson, D. R. (1956). Maximal flow through a network. *Canadian Journal of Mathematics*, 8:399–404. [13](#)
- [38] Fortin, P., Morency, C., and Trépanier, M. (2016). Innovative gtfs data application for transit network analysis using a graph-oriented method. *Journal of Public Transportation*, 19(4):2. [77](#)
- [39] Fortson, B. L., Klevens, J., Merrick, M. T., Gilbert, L. K., and Alexander, S. P. (2016). Preventing child abuse and neglect: A technical package for policy, norm, and programmatic activities. Technical report, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, Atlanta, GA. [39](#), [40](#), [45](#), [64](#)
- [40] Foulds, L. R. (2012). *Graph theory applications*. Springer Science & Business Media. [2](#)
- [41] Fulkerson, D. R. and Dantzig, G. B. (1955). Computation of maximal flows in networks. Technical report, RAND CORP SANTA MONICA CA. [12](#)

- [42] Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., and Colaneri, M. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*. 37
- [43] Global, C. V. (2020). Where we work. <https://cvg.org/where-we-work/>. 37
- [44] Guerra, F. M., Bolotin, S., Lim, G., Heffernan, J., Deeks, S. L., Li, Y., and Crowcroft, N. S. (2017). The basic reproduction number ($\mathcal{R}_0$) of measles: a systematic review. *Lancet Infectious Disease*, 17. 9
- [45] Gundersen, C. and Ziliak, J. P. (2015). Food insecurity and health outcomes. *Health affairs*, 34(11):1830–1839. 76
- [46] Heesterbeek, J. A. P. (2002). A brief history of r_0 and a recipe for its calculation. *Acta biotheoretica*, 50(3):189–204. 3, 9
- [47] Herrenkohl, T., Maguin, E., Hill, K., Hawkins, J., Abbott, R., and Catalano, R. (2000). Developmental risk factors for youth violence. *J Adolesc Health.*, (26(3):176-186). 39, 45, 64
- [48] HP, R. R. H. (1917). An application of the theory of probabilities to the study of a priori pathometry: part iii. *Proc. R. Soc. Lond. A*, 93:225–240. 3
- [49] Keeling, M. J. and Eames, K. T. (2005). Networks and epidemic models. *Journal of the Royal Society*, (2). 10
- [50] Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721. 3
- [51] Kirby, E. C., Mallion, R. B., Polla, P., and Skrzynski, P. J. (2016). What kirchhoff actually did concerning spanning trees in electrical networks and its relationship to modern graph-theoretical work. *Croatica Chemica Acta*, 89(4):1–16. 2
- [52] Koch, D., Illner, R., and Ma, J. (2013). Edge removal in random contact networks and the basic reproduction number. *Journal of mathematical biology*, 67(2):217–238. 10

- [53] Lawrence, C. o. (2021). Streets. <https://lawrenceks.org/maps/#gis-data>. 79
- [54] Li, J., Blakeley, D., and Smith?, R. J. (2011). The Failure of $\mathcal{R}_0$. *Computational and Mathematical Methods in Medicine*, 2011. 10
- [55] Liu, Y., Gayle, A. A., Wilder-Smith, A., and Rocklöv, J. (2020). The reproductive number of covid-19 is higher compared to sars coronavirus. *Journal of travel medicine*. 9
- [56] Lloyd, A. L., Kitron, U., Perkins, T. A., Vazquez-Prokopec, G. M., and Waller, L. A. (2020). The basic reproductive number for disease systems with multiple coupled heterogeneities. *Mathematical biosciences*, 321:108294. 9
- [57] Loosemore, B. (2019). How can cities end food deserts? here are 4 solutions that worked. 77
- [58] Lum, K. and Isaac, W. (2016). To predict and serve? *Significance Magazine*. 38
- [59] Madison, C. o. (2017). City of madison open data. <https://data-cityofmadison.opendata.arcgis.com>. 147, 148
- [60] Martcheva, M., Tuncer, N., and St Colette, M. (2015). Coupling Within-Host and Between-Host Infectious Diseases Models. *Biomath*, (4(2)). 38
- [61] Mitchell, C. and Kribs, C. (2017). A comparison of methods for calculating the basic reproductive number for periodic epidemic systems. *Bulletin of mathematical biology*, 79(8):1846–1869. 9, 10
- [62] National Research, C. (2009). National Academies Press (US) Copyright 2009 by the National Academy of Sciences. All rights reserved., Washington (DC). Review Book NBK208019 [bookaccession]. 76
- [63] NCDOT (2021). Road data. <https://connect.ncdot.gov/resources/gis/pages/gis-data-layers.aspx>. 79
- [64] Nurius, P. S., Russell, P. L., Herting, J. R., Hooven, C., and Thompson, E. A. (2009). Risk and protective profiles among never exposed, single form, and multiple form violence exposed youth. *J Child Adolesc Trauma*, (2(2): 106-123). 36, 39, 45, 64

- [65] Office, M. C. L. I. (2018). Street centerlines. <https://gis-mclio.opendata.arcgis.com/search?groupIds=46a9d715fee74e92af8d7fe9b8acf078>. 79
- [66] Ogier, R. G. (1988). Minimum-delay routing in continuous-time dynamic networks with piecewise-constant capacities. *Networks*, 18(4):303–318. 33
- [67] Philpott, A. B. (1990). Continuous-time flows in networks. *Mathematics of Operations Research*, 15(4):640–661. 33
- [68] Ploeg, M. V., Nulph, D., and Williams, R. (2011). Mapping food deserts in the united states. *Amber Waves*. 5, 76
- [69] Riaz, F. and Ali, K. M. (2011a). Applications of graph theory in computer science. In *2011 Third International Conference on Computational Intelligence, Communication Systems and Networks*, pages 142–145. IEEE. 2
- [70] Riaz, F. and Ali, K. M. (2011b). Applications of graph theory in computer science. In *2011 Third International Conference on Computational Intelligence, Communication Systems and Networks*, pages 142–145. IEEE. 2
- [71] Riedel, S. (2005). Edward jenner and the history of smallpox and vaccination. In *Baylor University Medical Center Proceedings*, volume 18, pages 21–25. Taylor & Francis. 1
- [72] Roberts, F. S. (1978). *Graph theory and its applications to problems of society*. SIAM. 10
- [73] Rodríguez, N. and Winkler, M. (2020). Relaxation by nonlinear diffusion enhancement in a two-dimensional cross-diffusion model for urban crime propagation. *Mathematical Models and Methods in Applied Sciences*, 30(11):2105–2137. 38
- [74] Ross, R. (1916). An application of the theory of probabilities to the study of a priori pathometry.—part i. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 92(638):204–230. 3

- [75] Ross, R. and Hudson, H. P. (1917). An application of the theory of probabilities to the study of a priori pathometry.—part ii. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 93(650):212–225. [3](#)
- [76] Ruan, S. (2017). Modeling the transmission dynamics and control of rabies in china. *Mathematical biosciences*, 286:65–93. [27](#), [28](#)
- [77] Saidi, S., Wirasinghe, S. C., Kattan, L., and Esmailnejad, S. (2017). A generalized framework for complex urban rail transit network analysis. *Transportmetrica A: Transport Science*, 13(10):874–892. doi: 10.1080/23249935.2017.1348401. [77](#)
- [78] Schichl, H. (2004). Models and the history of modeling. In *Modeling languages in mathematical optimization*, pages 25–36. Springer. [1](#)
- [79] Silk, M. J., Carrignon, S., Bentley, R. A., and Fefferman, N. H. (2022). Observations and conversations: how communities learn about infection risk can impact the success of non-pharmaceutical interventions against epidemics. *BMC public health*, 22(1):1–12. [37](#)
- [80] Skogan, W. G., Hartnett, S. M., Bump, N., and Dubois, J. (2008). Evaluation of ceasefire—chicago. Technical report, Northwestern University, Evanston, IL. [37](#)
- [81] Slutkin, G. (2013). Violence is a Contagious Disease. In *Contagion of Violence: Workshop Summary*. The National Academies Press. [4](#), [36](#), [37](#)
- [82] Smith, D., Moore, L., et al. (2004). The sir model for spread of disease-the differential equation model. *Convergence*. [3](#)
- [83] Smith, D. L., McKenzie, F. E., Snow, R. W., and Hay, S. I. (2007). Revisiting the basic reproductive number for malaria and its implications for malaria control. *PLoS biology*, 5(3):e42. [9](#)
- [84] Sooknanan, J., Bhatt, B., and Comissiong, D. M. (2013). Another way of thinking: a review of mathematical models of crime. *Math Today*, 131. [38](#)
- [85] Stewart, A. J. and Devlin, P. M. (2006). The history of the smallpox vaccine. *Journal of infection*, 52(5):329–334. [1](#)

- [86] Tse, W. and Ward, M. (2016). Hotspot formation and dynamics for a continuum model of urban crime. *European Journal of Applied Mathematics*, 27(3):583–624. 38
- [87] USDA (2021a). Documentation. <https://www.ers.usda.gov/data-products/food-environment-atlas/documentation/>. 79
- [88] USDA (2021b). Documentation. <https://www.ers.usda.gov/data-products/food-access-research-atlas/documentation/#definitions>. 94
- [89] USDA (2021c). Go to the atlas. <https://www.ers.usda.gov/data-products/food-access-research-atlas/go-to-the-atlas/>. 78
- [90] Valega-Mackenzie, W. and Ríos-Soto, K. (2018). Can Vaccination Save a Zika Virus Epidemic? *Bulletin of Mathematical Biology*, 80. 9
- [91] van den Driessche, P. and Watmough, J. (2002). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, (180). 3, 4, 9, 10, 13, 14
- [92] Wan, Z., Mahajan, Y., Kang, B. W., Moore, T. J., and Cho, J.-H. (2021). A survey on centrality metrics and their network resilience analysis. *IEEE Access*, 9:104773–104819. 82
- [93] Webster, D. W., Whitehill, J. M., Vernick, J. S., and Curriero, F. C. (2012). Effects of Baltimore’s Safe Streets Program on Gun Violence: A Replication of Chicago’s CeaseFire Program. *Journal of Urban Health*, (1). 36, 61, 63, 64
- [94] Weitz, J. S. and Dushoff, J. (2015). Modeling post-death transmission of ebola: challenges for inference and opportunities for control. *Scientific reports*, 5(1):1–7. 29, 30
- [95] Widener, M. J., Metcalf, S. S., and Bar-Yam, Y. (2013). Agent-based modeling of policies to improve urban food access for low-income populations. *Applied Geography*, 40:1–10. 77

- [96] Zipkin, J. R., Short, M. B., and Bertozzi, A. L. (2014). Cops on the dots in a mathematical model of urban crime and police response. *Discrete & Continuous Dynamical Systems-B*, 19(5):1479. [38](#)

Appendices

A Appendix for Chapter 4

This appendix consists of additional tables and figures for Chapter 4. Specifically, Section A.1 contains the additional tables for the sensitivity analysis of the representative addresses and Section A.2 contains the theoretical example of analyzing a food desert network with changes made to the transit system.

List of Tables and Figures in Appendix A

- Table 1: Representative Addresses for Brown Deer.....Page 122
- Table 2: Representative Addresses for Lawrence.....Page 122
- Table 3: Representative Addresses for Albuquerque.....Page 123
- Table 4: Representative Addresses for Charlotte.....Page 124
- Table 5: Representative Addresses for Raleigh.....Page 125
- Table 6: Overlap Centrality Results.....Page 126
- Figure 1: Full Network for Lawrence.....Page 127
- Figure 2: Full Network for Albuquerque.....Page 128
- Figure 3: Full Network for Charlotte.....Page 129
- Figure 4: Full Network for Raleigh.....Page 130
- Table 7: Degree Results for Brown Deer.....Page 131
- Table 8: Degree Results for Lawrence.....Page 131
- Table 9: Degree Results for Albuquerque.....Page 132
- Table 10: Degree Results for Charlotte.....Page 133
- Table 11: Degree Results for Raleigh.....Page 134
- Table 12: Closeness Results for Brown Deer.....Page 134

- Table 13: Closeness Results for Lawrence.....Page 135
- Table 14: Closeness Results for Albuquerque.....Page 135
- Table 15: Closeness Results for Charlotte.....Page 136
- Table 16: Closeness Results for Raleigh.....Page 137
- Table 17: Closeness Results for Brown Deer with Unverified Grocery Stores.....Page 137
- Table 18: Closeness Results for Lawrence with Unverified Grocery Stores.....Page 138
- Table 19: Closeness Results for Albuquerque with Unverified Grocery Stores.....Page 138
- Table 20: Closeness Results for Charlotte with Unverified Grocery Stores.....Page 139
- Table 21: Closeness Results for Raleigh with Unverified Grocery Stores.....Page 140

Section A.1

- Table 22: Representative Addresses for Brown Deer Selected with the Random Method.....Page 142
- Table 23: Representative Addresses for Brown Deer Selected with the Center Method.....Page 142
- Table 24: Representative Addresses for Lawrence Selected with the Random Method.....Page 143
- Table 25: Representative Addresses for Lawrence Selected with the Center Method.....Page 143
- Table 26: Degree Results for Brown Deer with Representative Addresses Selected Using the Random Method.....Page 143
- Table 27: Degree Results for Brown Deer with Representative Addresses Selected Using the Center Method.....Page 144

- Table 28: Degree Results for Lawrence with Representative Addresses Selected Using the Random Method.....Page 144
- Table 29: Degree Results for Lawrence with Representative Addresses Selected Using the Center Method.....Page 144
- Table 30: Closeness Results for Brown Deer with Representative Addresses Selected Using the Random Method.....Page 145
- Table 31: Closeness Results for Brown Deer with Representative Addresses Selected Using the Center Method.....Page 145
- Table 32: Closeness Results for Lawrence with Representative Addresses Selected Using the Random Method.....Page 145
- Table 33: Closeness Results for Lawrence with Representative Addresses Selected Using the Center Method.....Page 146

Section A.2

- Figure 5: Theoretical Food Desert Network.....Page 149
- Figure 6: Network with 10 Bus Stops Added Randomly.....Page 151
- Figure 7: Network with 5 Bus Stops Added Manually.....Page 152
- Figure 8: Network with 5 Random Edges Removed.....Page 153
- Table 34: Path Length Analysis for Original Network.....Page 154
- Table 35: Path Length Analysis for Network with Bus Stops Added Randomly.....Page 155
- Table 36: Path Length Analysis for Network with Bus Stops Added Manually.....Page 156
- Table 37: Path Length Analysis for Network with Edges Removed.....Page 157

Table 1: Representative Addresses for Brown Deer. Street addresses for the representative addresses in Brown Deer.

Food Desert	Address
101	Glenbrook Apartments (9151 N Jadam Ln)
800	Presido Square Apartments (5401 W Presidio Ln)
201	The Atrium Apartments (11601 W Brown Deer Rd)
302	Arbor Garden Apartments (10215 W Fond Du Lac Ave)
700	Jade Gardens Condominium (7849 N 60th St)
900	Teutonia Apartments (3744 W Green Tree Rd)

Table 2: Representative Addresses for Lawrence. Street addresses for the representative addresses in Lawrence.

Food Desert	Address
501	Northwinds Apartments (1311 George Ct)
002	Prairie Park Neighborhood (2524 Harper St)
100	North Lawrence Neighborhood (700 Lyon St)
801	Crossgate Casitas (2451 Crossgate Dr)
200	Maple Lane Apartments (1121 E 13th St)

Table 3: Representative Addresses for Albuquerque. Street addresses for the representative addresses in Albuquerque.

Food Desert	Address
000	Union 505 Apartments (801 Locust Pl)
736	Desert Willow Apartments (8901 Jefferson St NE)
707	Los Altos Towers (9125 Copper Ave NE)
500	Plaza David Chavez (2821 Mountain Rd NW)
102	Amherst Apartments (1111 Amherst Dr SE)
200	Lobo Village (1200 Avenida Cesar Chavez SE)
900	Cascade Crossing Apartments (3308 4th St NW)
901	Plaza Feliz Apartments (517 San Pablo St SE)
001	Paseo Del Rio (501 Poco Loco Dr SW)
734	Desert Springs Park Neighborhood (7627 Blue Avena Ave SW)
735	Tom Tenorio Park Neighborhood (2900 Arenal Rd SW)
738	Blossom Ridge HOA Park (9100-9198 Indigo Sky Tr SW)
401	South Atrisco Neighborhood (1201 Atrisco Rd SW)
604	Sunset Mobile Home Park (2305 Metzgar Rd SW)
713	Westgate Community Park (Valley View Dr SW)

Table 4: Representative Addresses for Charlotte. Street addresses for the representative addresses in Charlotte.

Food Desert	Address
510	Presley Oaks (4915 Misty Oaks Dr)
109	Trinity Park Apartments (9609 Trinity Rd)
403	Charlotte Spring Apartments (4825 Spring Trace Dr)
400	Firestone - Garden Park (3501 Firestone Dr)
500	Lakewood Neighborhood (3207 Kalynne St)
401	Northcross Townhomes Apartments (1835 Griers Grove Rd)
404	Woodstone Apartments (4826 Woodstone Dr)
006	AD Bennerman (9119 Spyglass Pl)
304	Colonial Place Apartments (1307 Springmont Ln)
303	Willow Oak Run Apartments (2908 Turning Oak Dr)
010	Coulwood Neighborhood (100 Coulwood Park Dr)
802	Pressley Ridge Apartments (1301 Pressley Rd)
201	Southgate Apartments (4001 Conway Ave l)
305	Tom Hunter Park (919 Tom Hunter Rd)
603	HillRock Estates (3317 Magnolia Hill Dr)
609	Valleyview Apartments (5117 Hickory Valley Ct)
616	Reedy Creek Neighborhood (4704 Meadowfield Rd)
620	Cedarwood Pointe Apartments (7700 Harrisburg Rd)
618	Steepleton Way Neighborhood (9164 Pleasant Ridge Rd)

Table 5: Representative Addresses for Raleigh. Street addresses for the representative addresses in Raleigh.

Food Desert	Address
713	Woodlake Downs Apartments (3151 Exacta Ln)
301	Clarion Crossing Apartments (1141 Crab Orchard Dr)
900	Junious N. Sorrell Park (416 E Lee St)
101	Cedar Crest Apartments (2821 Laodicea Dr)
500	Woodland Apartments (815 Suffolk Blvd # A)
803	Weston Trace Apartments (204 Westcroft Dr)
802	Legacy 2000 (2000 Spring Dr)
808	Abberly Place Apartments (316-320 Copperline Dr)
106	Village Lakes Neighborhood (5246 Turf Grass Ct)
809	King Author Trail Neighborhood (211 Lane of Sir Gallahad)
806	The Bluffs at Walnut Creek (2540 Bluffs View Dr)
018	River Birch at Town Center Apartments (2611 Torquay Crossing)
707	Autumn Trace Apartments (2401 Fall Ln)
112	Berkshire Park (1201 Aragon Drive)
108	Cottages at Knightdale Station (1100 Cannonball Run)
507	Enclave @ Crabtree (4009 Deep Hollow Dr)

Table 6: Overlap Centrality Results. Results from the overlap analysis. Grocery stores with an overlap of 1 are omitted from the chart.

Milwaukee		Albuquerque		Raleigh	
Store	Overlap	Store	Overlap	Store	Overlap
Save-A-Lot (6410 N 76th St)	2	Smith's Food & Drug (111 Coors Blvd NW)	3	Food Lion (1601 Cross Link Rd)	3
Pick'n Save (7401 W Good Hope Rd)	2	G & H Oriental Market (129 Pennsylvania St NE)	2	Save-A-Lot (1610 Martin Luther King Jr Blvd)	2
		Albertsons Market (1625 Rio Bravo Blvd SW, Ste 11)	2		
		Mercado Y Carniceria Mesquite (201 San Pedro Dr SE)	2		
		Walmart Supercenter (301 San Mateo Blvd SE)	2		
		LA Montanita Co-Op Food Market (3500 Central Ave SE)	3		
		Sprouts Farmers Market (5112 Lomas Blvd NE)	2		
		Albertsons Market (6600 Central Ave SW)	3		
		Talin Market World Food Fare (88 Louisiana Blvd SE)	2		
		Walmart Neighborhood Market (9600 Sage Rd SW)	2		

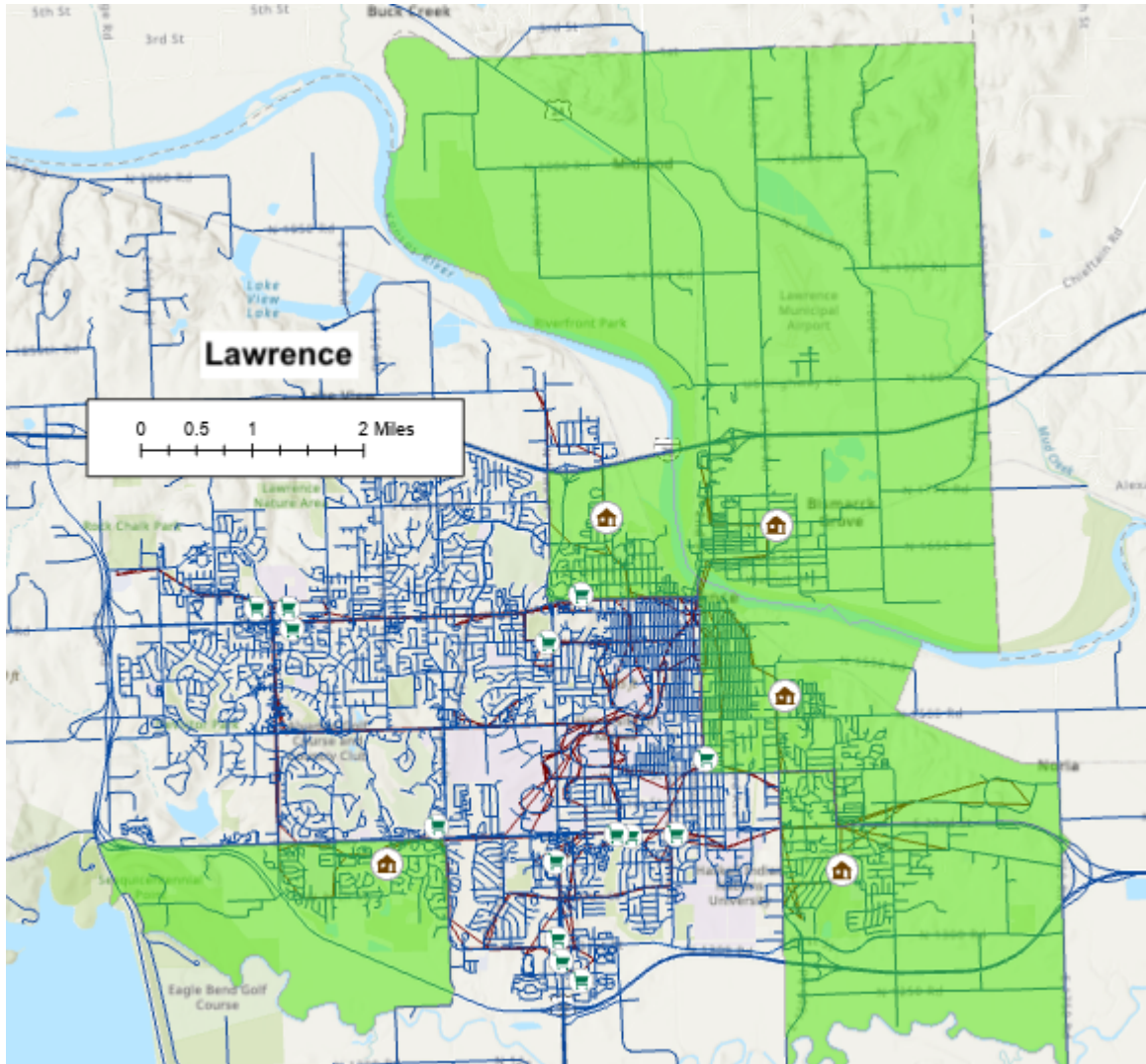


Figure 1: Full Network for Lawrence. Network model for Lawrence, KS. For the nodes, the food desert census tracts are shaded in green, the representative addresses are symbolized by a white circle with a picture of a house inside, and the grocery stores are symbolized by a white circle with a shopping cart inside. For the edges, the maroon lines represent the transit lines and the blue lines represent the streets associated to the transit system. Note, the maroon lines do not always follow the streets because they are representative of logical connections made by The Public Transit Evaluator tool to determine travel times in the transit system [31].

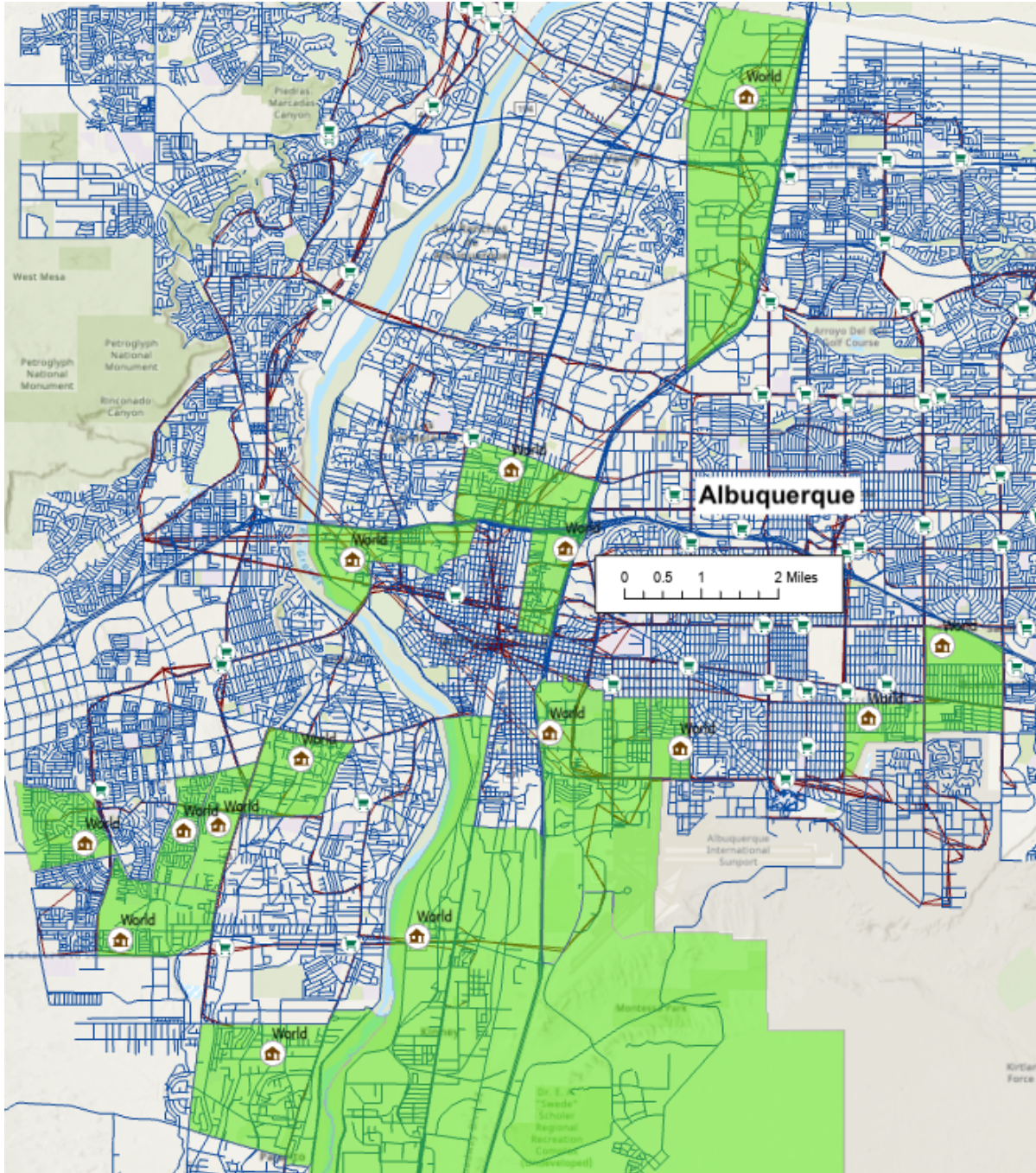


Figure 2: Full Network for Albuquerque. Network model for Albuquerque, NM. For the nodes, the food desert census tracts are shaded in green, the representative addresses are symbolized by a white circle with a picture of a house inside, and the grocery stores are symbolized by a white circle with a shopping cart inside. For the edges, the maroon lines represent the transit lines and the blue lines represent the streets associated to the transit system. Note, the maroon lines do not always follow the streets because they are representative of logical connections made by The Public Transit Evaluator tool to determine travel times in the transit system [31].

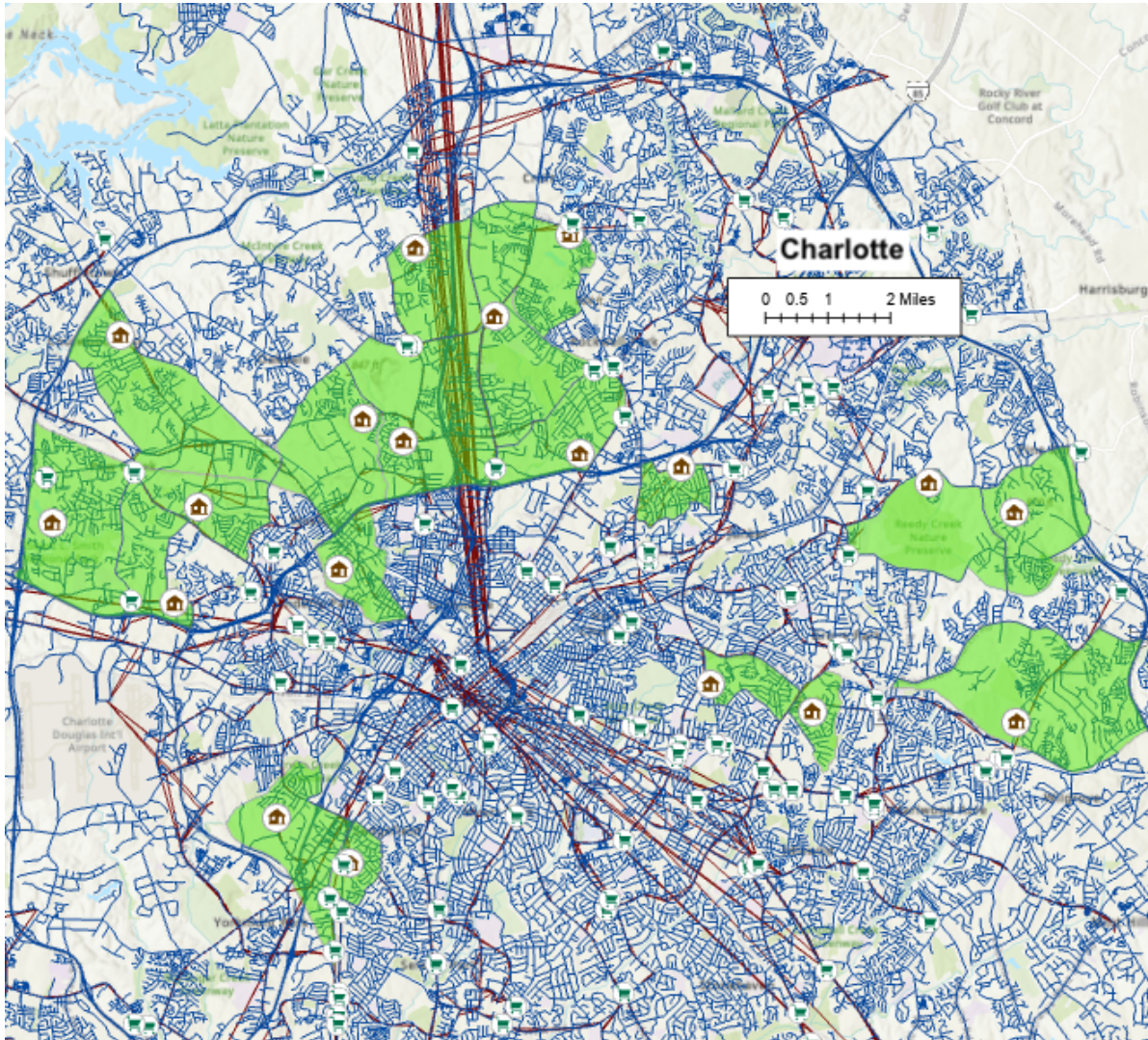


Figure 3: Full Network for Charlotte. Network model for Charlotte, NC. For the nodes, the food desert census tracts are shaded in green, the representative addresses are symbolized by a white circle with a picture of a house inside, and the grocery stores are symbolized by a white circle with a shopping cart inside. For the edges, the maroon lines represent the transit lines and the blue lines represent the streets associated to the transit system. Note, the maroon lines do not always follow the streets because they are representative of logical connections made by The Public Transit Evaluator tool to determine travel times in the transit system [31].

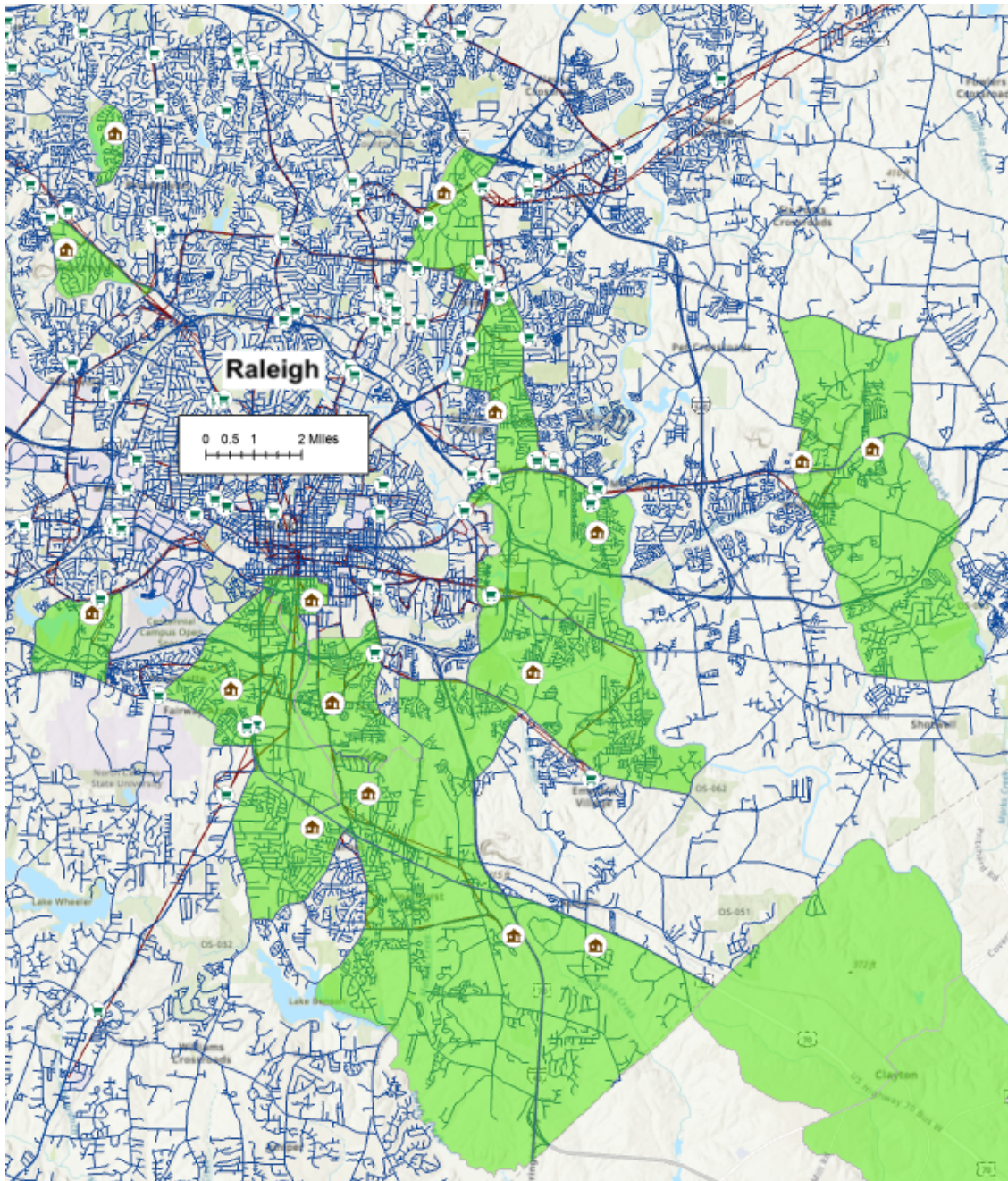


Figure 4: Full Network for Raleigh. Network model for Raleigh, NC. For the nodes, the food desert census tracts are shaded in green, the representative addresses are symbolized by a white circle with a picture of a house inside, and the grocery stores are symbolized by a white circle with a shopping cart inside. For the edges, the maroon lines represent the transit lines and the blue lines represent the streets associated to the transit system. Note, the maroon lines do not always follow the streets because they are representative of logical connections made by The Public Transit Evaluator tool to determine travel times in the transit system [31].

Table 7: Degree Results for Brown Deer. Number of bus stops within the stated distance from the representative address in the food desert.

Food Desert	Within 0.25 Miles	Within 0.5 Miles	Within 0.75 Miles
101	0	3	18
800	0	8	17
201	3	9	12
302	0	1	9
700	6	8	12
900	2	17	30

Table 8: Degree Results for Lawrence. Number of bus stops within the stated distance from the representative address in the food desert.

Food Desert	Within 0.25 Miles	Within 0.5 Miles	Within 0.75 Miles
501	2	8	14
002	4	7	20
100	2	7	14
801	4	10	17
200	4	8	22

Table 9: Degree Results for Albuquerque. Number of bus stops within the stated distance from the representative address in the food desert.

Food Desert	Within 0.25 Miles	Within 0.5 Miles	Within 0.75 Miles
000	1	7	41
736	10	21	25
707	4	17	39
500	0	4	17
102	6	17	29
200	0	8	27
900	10	31	60
901	2	24	52
001	0	2	5
734	0	6	17
735	1	9	15
738	0	8	12
401	2	18	31
604	0	0	5
713	0	16	30

Table 10: Degree Results for Charlotte. Number of bus stops within the stated distance from the representative address in the food desert.

Food Desert	Within 0.25 Miles	Within 0.5 Miles	Within 0.75 Miles
510	0	2	8
109	4	10	16
403	8	14	20
400	8	15	18
500	0	13	31
401	5	21	45
404	4	15	26
006	0	0	0
304	4	9	16
303	1	7	12
010	1	2	2
802	4	8	18
201	9	17	43
305	3	14	25
603	6	12	26
609	4	16	23
616	0	2	4
620	0	4	11
618	0	0	0

Table 11: Degree Results for Raleigh. Number of bus stops within the stated distance from the representative address in the food desert.

Food Desert	Within 0.25 Miles	Within 0.5 Miles	Within 0.75 Miles
713	0	0	0
301	3	13	24
900	12	43	73
101	8	25	40
500	5	13	25
803	0	5	11
802	0	0	9
808	2	6	8
106	0	0	2
809	0	0	0
806	0	0	2
018	2	5	16
707	2	5	7
112	2	5	15
108	0	0	2
507	3	12	19

Table 12: Closeness Results for Brown Deer. Number of grocery stores reachable by public transit and walking within the stated time from the representative address in the food desert.

Food Desert	Number of Stores within 0-10 minutes	Number of Stores within 10-20 minutes	Number of Stores within 20-30 minutes
302	0	0	1
201	0	0	0
700	0	0	0
900	1	1	5
800	0	0	1
101	0	0	2

Table 13: Closeness Results for Lawrence. Number of grocery stores reachable by public transit and walking within the stated time from the representative address in the food desert.

Food Desert	Number of Stores within 0-10 minutes	Number of Stores within 10-20 minutes	Number of Stores within 20-30 minutes
801	0	1	2
002	0	0	4
200	0	0	1
501	0	1	0
100	0	0	0

Table 14: Closeness Results for Albuquerque. Number of grocery stores reachable by public transit and walking within the stated time from the representative address in the food desert.

Food Desert	Number of Stores within 0-10 minutes	Number of Stores within 10-20 minutes	Number of Stores within 20-30 minutes
604	0	0	1
713	0	1	0
738	0	1	0
734	0	0	0
735	0	0	3
401	0	0	3
500	0	0	3
001	0	1	0
200	0	0	1
102	0	0	1
901	0	2	7
000	0	0	0
900	0	1	1
707	0	4	18
736	0	0	0

Table 15: Closeness Results for Charlotte. Number of grocery stores reachable by public transit and walking within the stated time from the representative address in the food desert.

Food Desert	Number of Stores within 0-10 minutes	Number of Stores within 10-20 minutes	Number of Stores within 20-30 minutes
802	0	0	0
201	1	2	1
603	0	0	5
609	0	0	1
620	0	2	0
006	0	0	0
303	0	1	1
304	0	0	0
010	0	0	0
500	0	0	1
400	0	0	0
401	0	0	1
404	0	0	1
403	0	0	1
109	0	0	0
510	0	0	2
305	0	2	0
616	0	0	1
618	0	0	0

Table 16: Closeness Results for Raleigh. Number of grocery stores reachable by public transit and walking within the stated time from the representative address in the food desert.

Census	Number of Stores within 0-10 minutes	Number of Stores within 10-20 minutes	Number of Stores within 20-30 minutes
802	0	0	0
101	1	0	1
301	1	0	0
500	0	1	3
900	0	1	3
808	0	0	0
809	0	0	0
803	0	0	0
806	0	0	0
707	0	0	6
106	0	2	0
112	0	0	0
108	0	0	0
507	0	1	3
713	0	0	2
018	0	3	1

Table 17: Closeness Results for Brown Deer with Unverified Grocery Stores. Number of grocery stores from the unverified list reachable by public transit and walking within the stated time from the representative address in the food desert.

Food Desert	Number of Stores within 0-10 minutes	Number of Stores within 10-20 minutes	Number of Stores within 20-30 minutes
302	0	0	1
201	0	1	0
700	0	0	0
900	1	8	16
800	0	0	1
101	0	1	4

Table 18: Closeness Results for Lawrence with Unverified Grocery Stores.
Number of grocery stores from the unverified list reachable by public transit and walking within the stated time from the representative address in the food desert.

Food Desert	Number of Stores within 0-10 minutes	Number of Stores within 10-20 minutes	Number of Stores within 20-30 minutes
801	0	1	2
002	0	0	7
200	0	0	2
501	0	1	1
100	0	0	0

Table 19: Closeness Results for Albuquerque with Unverified Grocery Stores.
Number of grocery stores from the unverified list reachable by public transit and walking within the stated time from the representative address in the food desert.

Food Desert	Number of Stores within 0-10 minutes	Number of Stores within 10-20 minutes	Number of Stores within 20-30 minutes
604	0	0	3
713	0	1	0
738	0	1	0
734	0	0	0
735	0	0	3
401	0	0	3
500	0	2	4
001	0	1	2
200	0	1	2
102	1	0	3
901	1	2	9
000	0	1	3
900	1	2	1
707	0	5	23
736	0	0	0

Table 20: Closeness Results for Charlotte with Unverified Grocery Stores.
Number of grocery stores from the unverified list reachable by public transit and walking within the stated time from the representative address in the food desert.

Food Desert	Number of Stores within 0-10 minutes	Number of Stores within 10-20 minutes	Number of Stores within 20-30 minutes
802	0	0	0
201	1	2	2
603	0	1	8
609	0	0	1
620	0	3	0
006	0	0	0
303	0	1	1
304	0	1	0
010	0	0	0
500	0	0	1
400	0	0	0
401	0	0	1
404	0	0	1
403	0	0	1
109	0	0	1
510	0	1	2
305	0	2	0
616	0	0	2
618	0	0	0

Table 21: Closeness Results for Raleigh with Unverified Grocery Stores. Number of grocery stores from the unverified list reachable by public transit and walking within the stated time from the representative address in the food desert.

Food Desert	Number of Stores within 0-10 minutes	Number of Stores within 10-20 minutes	Number of Stores within 20-30 minutes
802	0	0	0
101	2	3	1
301	1	0	0
500	0	2	4
900	2	3	6
808	0	0	0
809	0	0	0
803	0	0	0
806	0	0	0
707	0	1	8
106	0	2	0
112	0	0	0
108	0	0	0
507	0	1	4
713	0	0	2
018	0	3	2

A.1 Sensitivity Analysis of the Representative Addresses

This sections contains the tables with additional information for the sensitivity analysis of the representative addresses in Chapter [4](#).

Table 22: Representative Addresses for Brown Deer Selected with the Random Method. Street addresses for the representative addresses in Brown Deer selected by the random method.

Food Desert	Address
101	Harbor Pointe Apartments (9200 N 75th St)
800	Blancheand Madge Apartments (6965 N 55th St)
201	Garden Terrace Apartments (10851 W Donna Dr)
302	Maple Tree School (6644 N 107th St)
700	Faith Manor Apartments (7915 N 66th St)
900	Nathaniel Hawthorne School (6945 N 41st St)

Table 23: Representative Addresses for Brown Deer Selected with the Center Method. Street addresses for the representative addresses in Brown Deer selected by the center method.

Food Desert	Address
101	W Fairy Chasm Dr & N Northridge Mall Rd
800	4918 W Green Tree Rd
201	N 116th & W Brown Deer Rd
302	6735 N 109th St
700	N 67th and W Calumet Rd
900	N 38th & W Juniper Ct

Table 24: Representative Addresses for Lawrence Selected with the Random Method. Street addresses for the representative addresses in Lawrence selected by the random method.

Food Desert	Address
501	Holiday Apartments (211 Mount Hope Ct)
002	Laurel Glen Apartments (1401 E 24th St)
100	Woodlawn Elementary (508 Elm St)
801	Parkway Commons Apartments (3601 Clinton Pkway)
200	New York Elementary (936 New York St)

Table 25: Representative Addresses for Lawrence Selected with the Center Method. Street addresses for the representative addresses in Lawrence selected by the center method.

Food Desert	Address
501	100 Arkansas St
002	1623 N 1320 Rd
100	1588 N 1900 Rd
801	2705 Bluestem Ct
200	1622 Irving Ct

Table 26: Degree Results for Brown Deer with Representative Addresses Selected Using the Random Method. Number of bus stops within the stated distance from the representative addresses selected with the random method in the food deserts.

Food Desert	Within 0.25 Miles	Within 0.5 Miles	Within 0.75 Miles
101	5	8	19
800	0	9	14
201	6	12	21
302	2	8	13
700	0	8	16
900	4	18	28

Table 27: Degree Results for Brown Deer with Representative Addresses Selected Using the Center Method. Number of bus stops within the stated distance from the representative addresses selected with the center method in the food deserts.

Food Desert	Within 0.25 Miles	Within 0.5 Miles	Within 0.75 Miles
101	5	9	16
800	0	6	23
201	0	1	9
302	2	6	23
700	0	6	20
900	11	18	26

Table 28: Degree Results for Lawrence with Representative Addresses Selected Using the Random Method. Number of bus stops within the stated distance from the representative addresses selected with the random method in the food deserts.

Food Desert	Within 0.25 Miles	Within 0.5 Miles	Within 0.75 Miles
501	1	10	19
002	3	17	29
100	4	16	28
801	6	9	15
200	10	27	41

Table 29: Degree Results for Lawrence with Representative Addresses Selected Using the Center Method. Number of bus stops within the stated distance from the representative addresses selected with the random center in the food deserts.

Food Desert	Within 0.25 Miles	Within 0.5 Miles	Within 0.75 Miles
501	4	8	21
002	0	0	2
100	0	0	0
801	2	6	11
200	0	6	14

Table 30: Closeness Results for Brown Deer with Representative Addresses Selected Using the Random Method. Number of grocery stores reachable by public transit and walking within the stated time from the representative address selected with the random method in the food desert.

Food Desert	Number of Stores within 0-10 minutes	Number of Stores within 10-20 minutes	Number of Stores within 20-30 minutes
101	0	0	0
800	0	0	1
201	0	0	0
302	0	0	0
700	0	0	0
900	0	0	1

Table 31: Closeness Results for Brown Deer with Representative Addresses Selected Using the Center Method. Number of grocery stores reachable by public transit and walking within the stated time from the representative address selected with the center method in the food desert.

Food Desert	Number of Stores within 0-10 minutes	Number of Stores within 10-20 minutes	Number of Stores within 20-30 minutes
101	0	0	0
800	0	0	0
201	0	0	0
302	0	0	0
700	0	0	1
900	0	0	1

Table 32: Closeness Results for Lawrence with Representative Addresses Selected Using the Random Method. Number of grocery stores reachable by public transit and walking within the stated time from the representative address selected with the random method in the food desert.

Food Desert	Number of Stores within 0-10 minutes	Number of Stores within 10-20 minutes	Number of Stores within 20-30 minutes
501	0	1	1
002	0	0	1
100	0	0	0
801	1	0	1
200	0	0	1

Table 33: Closeness Results for Lawrence with Representative Addresses Selected Using the Center Method. Number of grocery stores reachable by public transit and walking within the stated time from the representative address selected with the center method in the food desert.

Food Desert	Number of Stores within 0-10 minutes	Number of Stores within 10-20 minutes	Number of Stores within 20-30 minutes
501	0	1	1
002	0	0	0
100	0	0	0
801	0	0	0
200	0	0	0

A.2 Path Length Analysis for a Theoretical Food Desert Network

We now explore a food desert network by path length analysis and determine how this analysis can be used on the city level to evaluate the role public transit plays in connecting food deserts to grocery stores. Specifically, we formulate a theoretical food desert network and look at how path length analysis changes when nodes and edges are added or removed, which represents changes in the transit system. By using a theoretical network, instead of one reflecting a real city, the complexity of the network and data-set decreases and allows us to focus on the types of results and interpretation that can come from this analysis rather than the specific implications for one location. In order for our theoretical network to have a realistic geographic distribution of food deserts, grocery stores, bus stops and bus lines, we loosely based our network on a portion of a real city (Madison, WI) without ensuring accuracy or interpreting our outcomes in the context of the region. To get the geographic layout we start by building a food desert map of Madison in ArcGIS Pro and then use the coordinate information of the food deserts, grocery stores, and bus stops to define the network and perform the analysis in MATLAB. Note, in MATLAB a network is called a graph, so graph and network will be used interchangeably from here on.

We start by using ArcGIS Pro to construct a map of the food deserts and public transit system (bus stops and lines) of Madison, WI using the same food desert data from the USDA as before [1] and data from the City of Madison’s open data base [59]. Then using the Reference USA/Data Axle Reference Solutions database we generate a list of grocery stores based on the SIC codes in Table 4.1 and plot them on the map along with points in the center of each food desert. Next we define our graph in MATLAB by importing the x, y coordinates of the grocery stores and food desert points as nodes into MATLAB. Note, ArcGIS Pro uses the WGS84 Web Mercator coordinate system, so the scale is small since we are only working with a small fraction of a world map. In addition to the grocery store and food desert nodes, this graph will have a third type of node: bus stops. To define the bus stop nodes we abstract and simplify the map of Madison’s transportation system. We do this by selecting x, y coordinates on the map to represent the bus stops that loosely follow

where the actual bus stops are in Madison [59] and then import those to MATLAB as nodes in the graph.

Next we add edges to our graph representing the bus lines. First, we manually define the edges between the bus stops nodes such that they loosely follow the shape of the real bus lines in Madison. Then we connect the food desert and grocery store nodes to the network by adding edges between them and the nearby bus stops. The edges between the bus stop nodes and food desert/grocery store nodes represent the distance an individual has to walk, while the edges between bus stop nodes represent distance traveled on public transit. Thus, the former edges are much shorter than the latter. The final piece of the network we need to define is the edge weighting. The edge weights are assigned by the Euclidean distance between the points representing the originating and terminating node as defined by the coordinate system used by ArcGIS Pro. The numerical values are calculated using MATLAB's `norm` function. Due to the small scale that came from ArcGIS Pro we multiplied the edge weight by 1000 so that the edge weights, and therefore the path lengths, would be near or greater than 1. This was done to save space and increase readability. The path length and distance between the nodes are measured spatially, rather than temporally as is the case for the rest of Chapter 4. With the edge weights defined, the network is complete and can be seen in Figure 5. It has 67 nodes in total (3 food deserts, 13 grocery stores, 51 bus stops) and 80 edges.

To analyze our network we will use path length analysis, which looks at the path that defines the shortest distances between two nodes based on the weights of the edges. We are specifically interested in the path lengths between the food desert nodes and all 13 grocery store nodes, and how it changes if new bus stop nodes and edges are added or removed. Our metric for the utility of these modifications will be the change (reduction or increase) in the Euclidean distance, d , between the points, as defined above. To calculate path length we use MATLAB's `distances` function, which finds the shortest path between two nodes based on the edge weights. We start by finding the path lengths between food deserts and grocery stores of the original network defined above and then look at what happens to the path

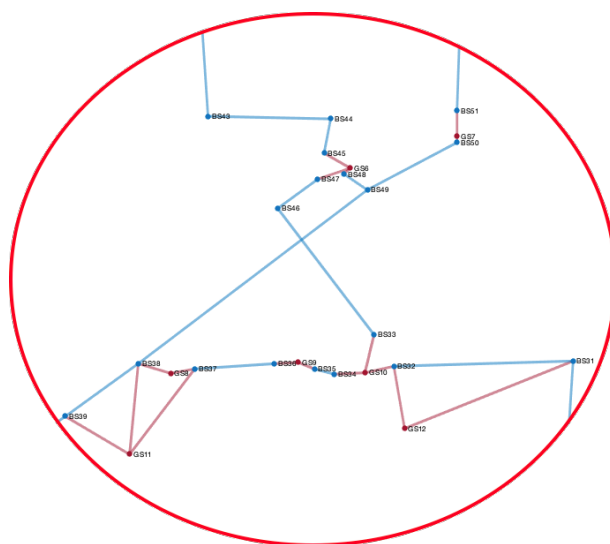
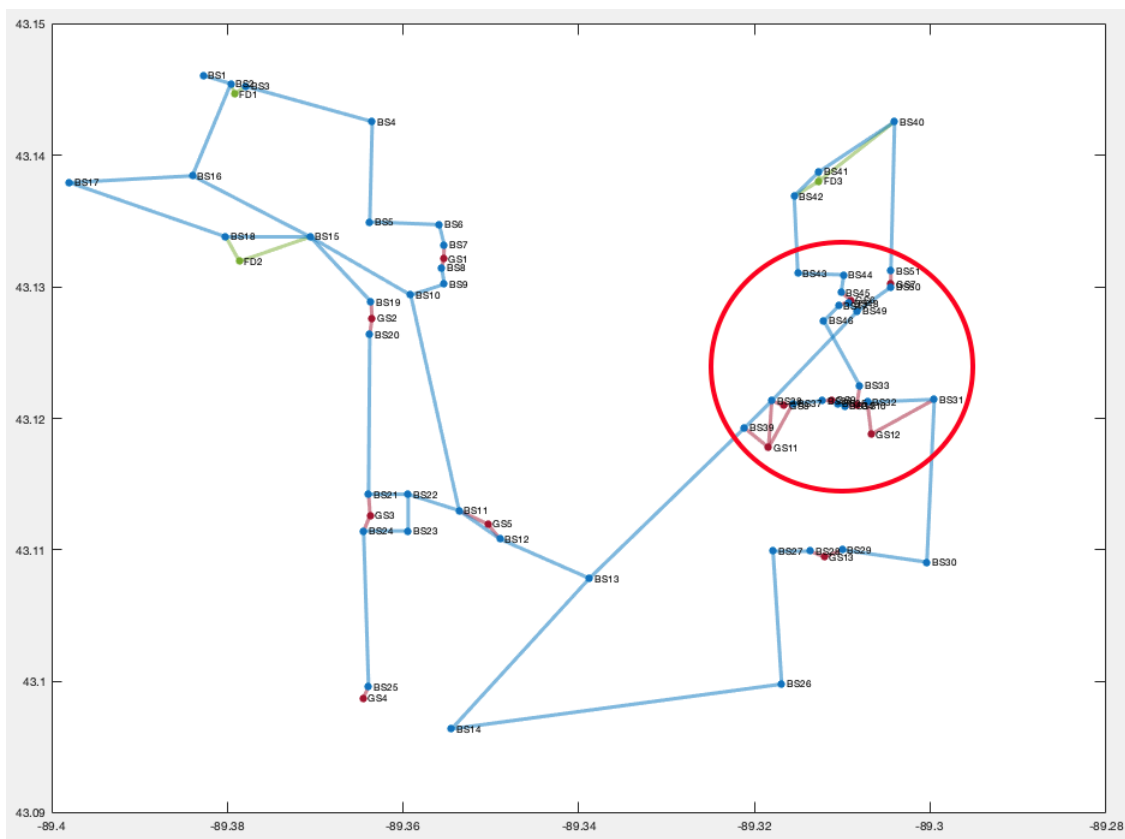
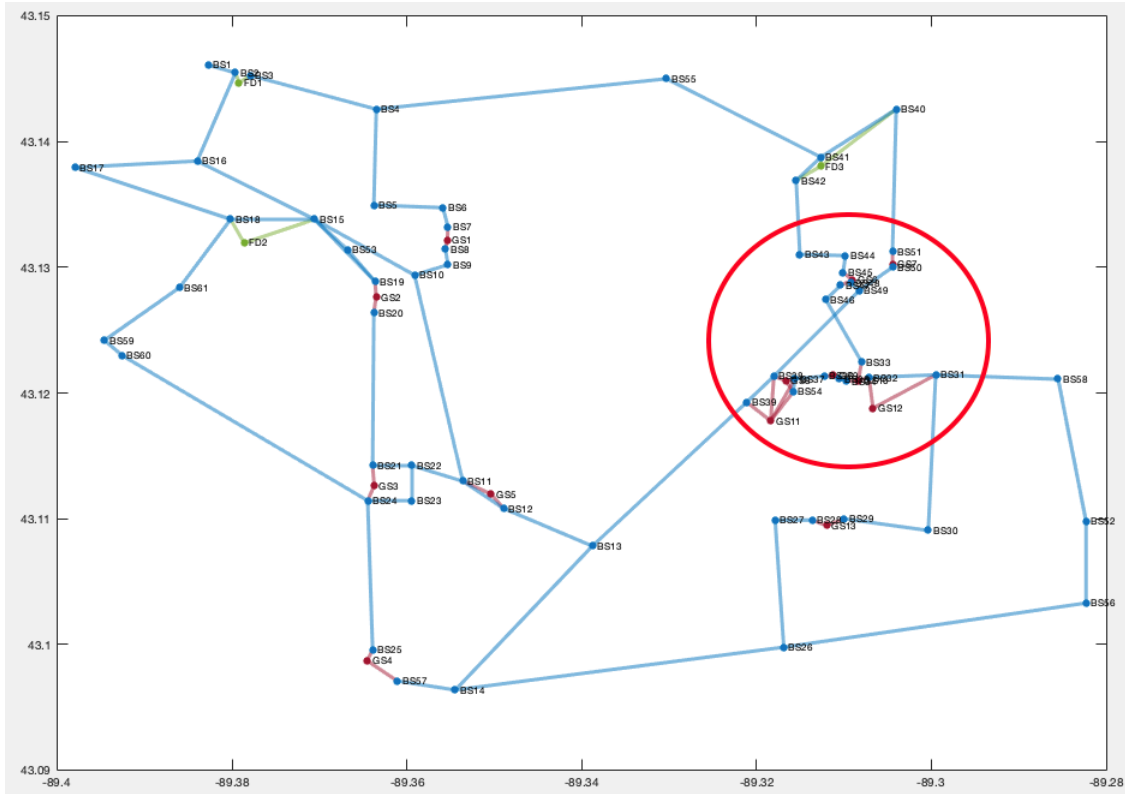


Figure 5: Theoretical Food Desert Network. The graphical representation of the food desert network. The food desert nodes, labeled FD1, FD2, FD3, are green, the grocery store nodes, labeled GD1, GS2..., are red, and the bus stop nodes, labeled BS1, BS2,..., are blue. The edges are colored based on their originating node.

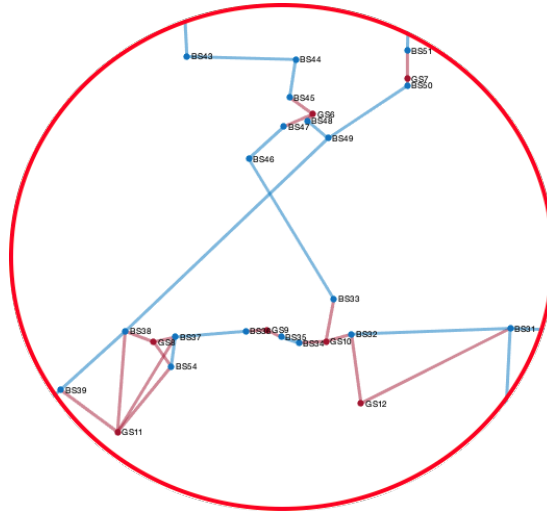
lengths in the following three scenarios: randomly adding 10 bus stops to the network, manually adding 5 bus stops, and removing 5 random edges from the network. We added the 10 random bus stop nodes to the network by using a random number generator on MATLAB to define 10 x, y coordinate pairs. After adding the new nodes to the graph, we manually added edges to connect them to the network. The edges were added to connect the bus stops to other near-by nodes in a manner that fit in with the the original edges. Figure 6 shows the resulting network. For the next scenario we looked at the original network and manually selected 5 locations for new bus stop nodes with the goal of decreasing the path lengths. This network is seen in Figure 7. We again added the edges in the same manner as before. The last scenario we considered is removing edges. We used a random number generator to select 5 edges that were then removed from the graph before calculating the path lengths. Figure 8 shows this case.

Tables 34-37 show the results of the path length analysis for the four scenarios. We compare the differences in the path length analyses for Tables 35-37 to Table 34. Table 35 shows that randomly adding 10 bus stops to the network decreases the distance to 6 grocery stores for Food Desert 1, 1 grocery store for Food Desert 2, and 2 grocery stores for Food Desert 3. Across all 3 food deserts the total reduction in d is 109.6765. In Table 36 we see that the distance between Food Desert 1 and all the grocery stores is decreased, for Food Desert 2 the distance to 1 grocery store is decreased, and for Food Desert 3 the distance to 6 grocery store is decreased when 5 bus stops are added manually. The overall reduction in d for all 3 food deserts is 418.0107. Table 37 shows when 5 random edges are removed d is increased by 6.7528 between Food Desert 3 and Grocery Store 5.

These four scenarios were chosen to approximate potential real world circumstances. The original network represents the current relationship between public transportation and food access. The other three scenarios look at what can happen when changes are made to the transit system by city officials. The network with 10 randomly added bus stop represents a situation in which bus stops are added near food deserts and grocery stores without explicitly

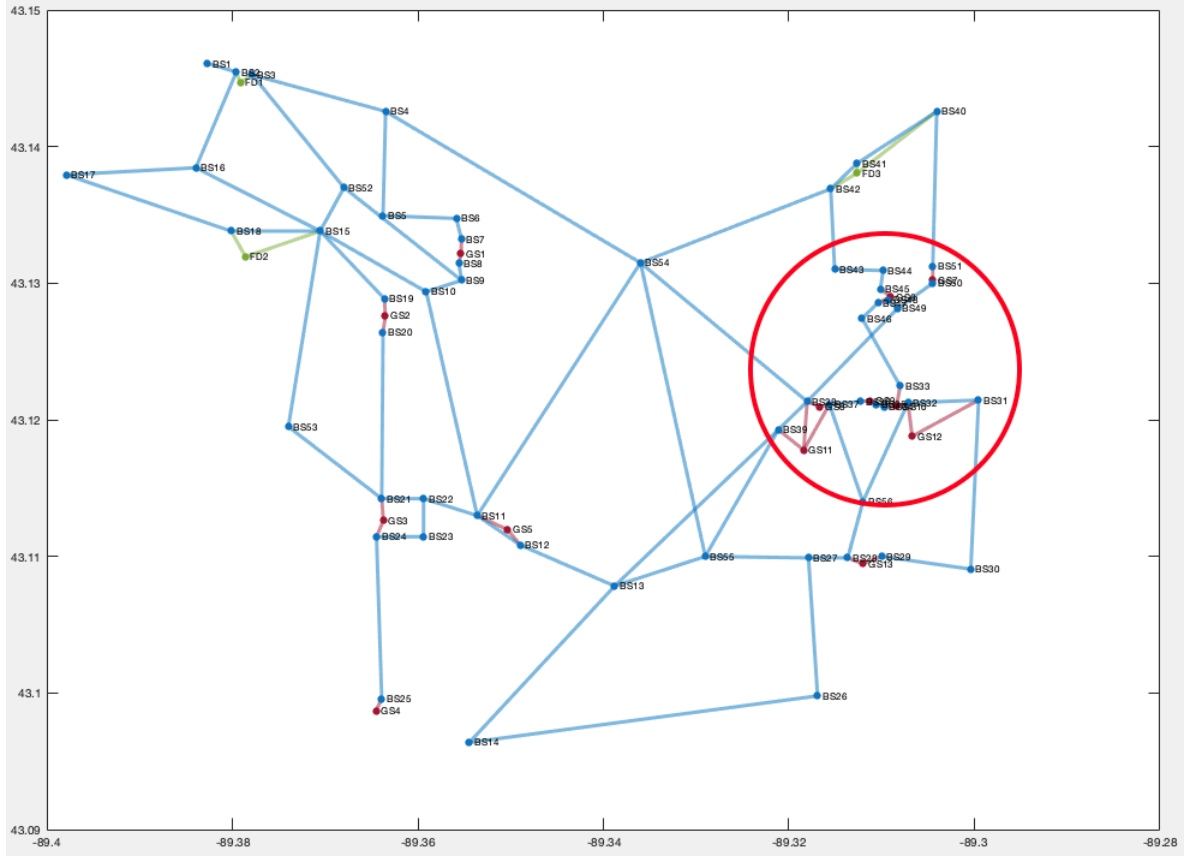


(a) Full network.

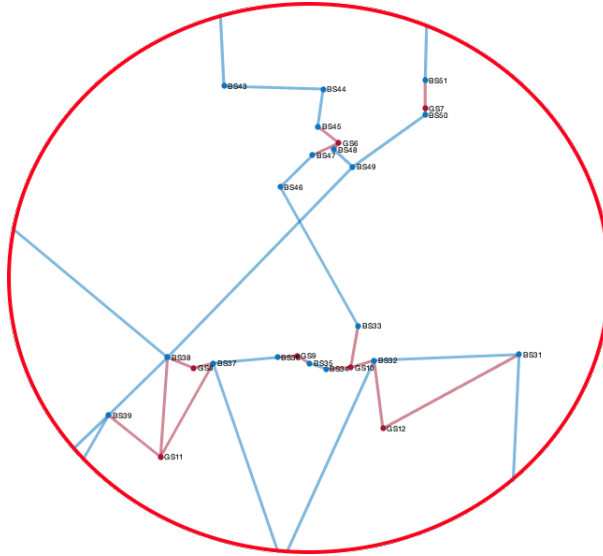


(b) Zoomed in to see detail.

Figure 6: Network with 10 Bus Stops Added Randomly. The graphical representation of the network with 10 bus stops added randomly. The food desert nodes, labeled FD1, FD2, FD3, are green, the grocery store nodes, labeled GD1, GS2..., are red, and the bus stop nodes, labeled BS1, BS2..., are blue. The edges are colored based on their originating node.



(a) Full network.



(b) Zoomed in to see detail.

Figure 7: Network with 5 Bus Stops Added Manually. The graphical representation of the network with 5 bus stops added manually. The food desert nodes, labeled FD1, FD2, FD3, are green, the grocery store nodes, labeled GD1, GS2..., are red, and the bus stop nodes, labeled BS1, BS2..., are blue. The edges are colored based on their originating node.

Table 34: Path Length Analysis for Original Network. The Euclidean distance based on the shortest path between each food desert and each grocery store for the original network is given.

	Food Desert 1	Food Desert 2	Food Desert 3
Grocery Store 1	34.2965	26.4640	93.9259
Grocery Store 2	33.1002	18.0443	94.5873
Grocery Store 3	48.1409	33.0851	82.7833
Grocery Store 4	62.4855	47.4296	97.1279
Grocery Store 5	56.3249	41.2690	67.3875
Grocery Store 6	106.8228	91.7669	16.8059
Grocery Store 7	109.7899	94.7341	22.0620
Grocery Store 8	94.8477	79.7918	31.6330
Grocery Store 9	100.3064	85.2506	31.1686
Grocery Store 10	103.2150	88.1591	28.2601
Grocery Store 11	92.8115	77.7557	33.7783
Grocery Store 12	106.9482	91.8924	31.9934
Grocery Store 13	136.2647	121.2088	61.3098

Table 35: Path Length Analysis for Network with Bus Stops Added Randomly.
The Euclidean distance based on the shortest path between each food desert and each grocery store is given. Numbers in green represents a decrease in distance from Table 34.

	Food Desert 1	Food Desert 2	Food Desert 3
Grocery Store 1	34.2965	26.4640	70.9673
Grocery Store 2	33.1002	18.0443	94.5873
Grocery Store 3	48.1409	33.0851	82.7833
Grocery Store 4	62.4855	47.4296	84.8777
Grocery Store 5	56.3249	41.2690	67.3875
Grocery Store 6	85.1854	91.7669	16.8059
Grocery Store 7	89.8351	94.7341	22.0620
Grocery Store 8	94.8477	79.7918	31.6330
Grocery Store 9	99.5481	85.2506	31.1686
Grocery Store 10	96.6396	88.1591	28.2601
Grocery Store 11	92.8115	77.7557	33.7783
Grocery Store 12	100.3729	91.8924	31.9934
Grocery Store 13	126.7815	111.7256	61.3098

Table 36: Path Length Analysis for Network with Bus Stops Added Manually.
The Euclidean distance based on the shortest path between each food desert and each grocery store is given. Numbers in green represents a decrease in distance from Table 34.

	Food Desert 1	Food Desert 2	Food Desert 3
Grocery Store 1	30.6989	26.4640	72.1866
Grocery Store 2	28.1447	18.0443	73.7119
Grocery Store 3	43.1854	33.0851	61.9079
Grocery Store 4	57.5300	47.4296	76.2525
Grocery Store 5	51.3694	41.2690	53.2648
Grocery Store 6	79.7597	91.7669	16.8059
Grocery Store 7	82.7268	94.7341	22.0620
Grocery Store 8	67.7846	79.7918	31.6330
Grocery Store 9	73.2433	85.2506	31.1686
Grocery Store 10	76.1519	88.1591	28.2601
Grocery Store 11	69.9299	77.7557	33.7783
Grocery Store 12	79.8852	91.8924	31.9934
Grocery Store 13	82.9315	80.6652	44.3436

Table 37: Path Length Analysis for Network with Edges Removed. The Euclidean distance based on the shortest path between each food desert and each grocery store is given. Numbers in red represents a increase in distance from Table 34.

	Food Desert 1	Food Desert 2	Food Desert 3
Grocery Store 1	34.2965	26.4640	93.9259
Grocery Store 2	33.1002	18.0443	94.5873
Grocery Store 3	48.1409	33.0851	82.7833
Grocery Store 4	62.4855	47.4296	97.1279
Grocery Store 5	56.3249	41.2690	74.1402
Grocery Store 6	106.8228	91.7669	16.8059
Grocery Store 7	109.7899	94.7341	22.0620
Grocery Store 8	94.8477	79.7918	31.6330
Grocery Store 9	100.3064	85.2506	31.1686
Grocery Store 10	103.2150	88.1591	28.2601
Grocery Store 11	92.8115	77.7557	33.7783
Grocery Store 12	106.9482	91.8924	31.9934
Grocery Store 13	136.2647	121.2088	61.3098

considering food access. The network with 5 manually added bus stops shows the results of modifying the transit system with the specific goal of increasing food access and the network with 5 random edges removed represents what can happen to food access when bus lines/stops are removed near food deserts and grocery stores without consideration of the impact of the action on food access.

We decided to look at adding more bus stops in two ways to see the difference between when food access is explicitly considered versus when it is not. We chose to add 10 bus stops in randomly and only 5 in manually as it could be more costly for a city to modify its transit system solely to focus on food access. This also allows us to see the difference in quality of bus stop placement over quantity. Unsurprisingly, when increasing food access is the explicit reason for adding bus stops and lines we see a more dramatic increase in food access. The overall distance between food deserts and grocery stores decreases approximately 4 times more with 5 manually placed bus stops than 10 randomly placed bus stops. Therefore, we can conclude that a few carefully placed bus stops/lines is more useful for increasing food access, than a larger amount that do not take food access into consideration.

While not the most effective way to alleviate food deserts, Table 35 does show that just increasing the amount of bus stops/lines, near food deserts has the potential to increase food access, which may be more practical for transit planners. Thus, any expansion of their transit system near food deserts and grocery stores, whether the motivation is food access or not, will most likely have a positive effect on food access. Although in this same vein, Table 37 shows that removing bus stops/lines near food deserts or grocery stores, without considering the effects it could have on food access, can negatively impact it. Thus, food access should be considered when modifying, especially eliminating, public transit. Ideally, any modifications to the transit system should actively try to increase food access when possible, but at the minimum it should be considered to ensure changes do not accidentally make access worse. These results support what we found in the rest of Chapter 4.

Vita

Anna Sisk was born in Blue Springs, Missouri to Darrin and Nicole Sisk. She spent the first half of her childhood in Lee's Summit, Missouri before moving with her family to Belvidere, Illinois. She began attending her local community college, Rock Valley College (RVC), during high school and received her Associates in Science in 2015. While at RVC, Anna's calculus I professor, Dr. Paul Gunsul, first sparked her interest in mathematics. She then transferred to Marquette University and received a Bachelor of Science degree in Mathematics with a minor in Biological Sciences in 2017. After participating in a Research Experience for Undergraduates at Marquette, she knew she wanted to continue doing research and decided to attend graduate school. She chose to attend the University of Tennessee, Knoxville to pursue a Doctor of Philosophy degree in Mathematics with a concentration in Mathematical Biology under Dr. Nina Fefferman. In 2019 she received her Masters of Science and in 2022 she graduated with her PhD. Her research interests include: mathematical modeling of real world systems (specifically human biology and disease, epidemiology, and social issues), dynamical systems, and network theory. After graduation, she plans to pursue a career in research. When she is not doing math, Anna enjoys going on walks and hikes with her dog, Gauss, and fiancé, Wencel. She is incredibly grateful for all the love and support from her family, friends, Wencel, and Gauss through her education and as she begins her new career.