UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

PROTECTING INFRASTRUCTURE NETWORKS FROM DISINFORMATION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

SAEED JAMALZADEH
Norman, Oklahoma
2023

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/337460

PROTECTING INFRASTRUCTURE NETWORKS FROM DISINFORMATION

A DISSERTATION APPROVED FOR THE
SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING

BY THE COMMITTEE CONSISTING OF

Prof. Kash Barker (Chair)

Prof. Andrés D. González

Prof. Sridhar Radhakrishnan

Prof. Talayeh Razzaghi

Prof. Yifu Li

© Copyright by SAEED JAMALZADEH 2023
All Rights Reserved.

Acknowledgments

I would like to thank my advisor, Professor Kash Barker, who has made a great impact in my personal and professional life with his wisdom. Without his leadership the creation of this dissertation would not have been possible. Many thanks to Professors Andrés D. González and Sridhar Radhakrishnan for enriching my work with their knowledge and expertise. I also wish to thank Professors Talayeh Razzaghi and Yifu Li for serving on my dissertation committee.

I am extremely grateful to Professors Elena Bessarabova, Jonas Johansson, and Giovanni Sansavini for their contribution and the ideas they brought to this dissertation. I would also like to express my gratitude to Lily Mettenbrink. I am impressed by her research abilities, which made an extensive improvement on this dissertation.

I wish to thank my loving wife, Nasim, her and my wonderful family who provide eternal love and make our world more peaceful. Last but not least, my greatest appreciation is for the Sooner faculty and staff members for their unforgettable hospitality.

Table of Contents

Acknowledgments **iv**

List of Tables **vii**

List of Figures **viii**

Abstract **x**

1 Introduction **1**
1.1 Integrated Information-Infrastructure Networks 1
1.2 Models and Formulations 2
1.3 Structure of the Dissertation 5

2 Protecting Infrastructure Performance from Disinformation Attacks **7**
2.1 Introduction and Motivation 7
2.2 Background and Literature Review 9
2.2.1 Models of Information and Disinformation Spread 9
2.2.2 Models of Infrastructure Flow Optimization 12
2.3 Proposed Integrated Epidemiological + Optimization (EPO) Model . . 14
2.3.1 SIR Model 15
2.3.2 Network Flow Balance Optimization Model 18
2.4 Case Study: Performance of an Electric Power Network under Disinformation Attack 22
2.4.1 Determining the Parameters of the Model 24
2.4.2 Numerical Results 27
2.5 Concluding Remarks 36

3 Interdicting Disinformation to Prevent Infrastructure Disruption **39**
3.1 Introduction and Motivation 39
3.2 Background and Literature Review 42
3.3 Proposed Disinformation Interdiction Model 46
3.4 Case Study: Subway Network Performance Under a Weaponized Disinformation Attack 55
3.4.1 Data and Parameters of the Proposed Model 58
3.4.2 Numerical Results 59
3.5 Concluding Remarks 68

4 Dealing with Uncertainty in Weaponized Disinformation Attacks and Infrastructure Disruption **70**
4.1 Introduction 70
4.2 Background and Literature Review 74
4.3 Proposed Robust Model for Network Protection under Uncertain Disinformation Propagation 78
4.3.1 Deterministic Mixed Integer Programming Model 78
4.3.2 Robust Mixed Integer Programming Model with Polytopic Uncertainty Set 81
4.4 Case Study: Robust Protection of the Electric Power Network in Disinformation Campaigns 88
4.4.1 Data and Model Parameters 90
4.4.2 Numerical Results 93
4.5 Concluding Remarks 98

5 Conclusions **101**
5.1 Summary of Conclusions 101
5.2 Future Directions 103

Reference List **106**

List of Tables

2.1 Model notation. 21
2.2 Parameter values. 33

3.1 Optimization model notation. 50

4.1 Model notation. 79

List of Figures

1.1 Interdependent information and physical infrastructure networks, where disinformed users shown by red nodes in the information network can disrupt nodes in the physical infrastructure network. 3

2.1 Distribution of electric power demand and population, LA County, USA. 29
2.2 The results of (a) average infected communities, (b) average shortage, (c) average flows, and (d) average counter information targets with different combinations of β and γ as percentage of the baseline value. 33
2.3 Time series of (a) average infected communities, (b) average shortage, (c) average flow, and (d) average counter information targets, with different combinations of β and γ. Note the relationship to particular points in Figure 2.2 denoted by shape. 36

3.1 New York City Subway network. 61
3.2 Social interaction graph before and after interdiction under the scenario morning split-up. 62
3.3 (a) A route not impacted by disinformation; (b) A route impacted by disinformation (reroute), both under the split-up scenario. 64
3.4 Percentage of delay improvement (reduction) under scenarios (a) split-up in the morning, (b) split-up in the afternoon, (c) link-up in the morning, and (d) link-up in the afternoon. 65
3.5 (a) The first 5 stations closed by rumor under the split-up scenario; (b) The first 12 stations closed by rumor under the split-up scenario. . . . 66
3.6 Three stations most impacted by interdicting disinformation under the link-up scenario in the morning. 67

4.1 Los Angeles County electric power network. 92
4.2 The percentage of improvement in the optimal solution by solving the robust optimization model relative to the deterministic optimization problem. 94
4.3 Optimal improvement of the solution by solving the robust optimization model compared to the deterministic model with (a) fixed rate of user-adopted disinformation and unfixed rate of disinformation detection, (b) fixed rate of disinformation detection and unfixed rate of user-adopted disinformation, and (c) unfixed rates of user-adopted disinformation and disinformation detection, for the combination of parameters $\beta = 1, \gamma = \frac{1}{5}$. 96

4.4 Change in the average counter information targets by solving the robust optimization model compared to the deterministic model with (a) fixed rate of user-adopted disinformation and unfixed rate of disinformation detection, (b) fixed rate of disinformation detection and unfixed rate of user-adopted disinformation, and (c) unfixed rates of user-adopted disinformation and disinformation detection. 97
4.5 Optimal solutions suggested by solving the robust optimization model compared to the deterministic model with (a) fixed rate of user-adopted disinformation and unfixed rate of disinformation detection, (b) fixed rate of disinformation detection and unfixed rate of user-adopted disinformation, and (c) unfixed rates of user-adopted disinformation and disinformation detection. 98

Abstract

Massive amount of information shared on online platforms makes the verification of contents time-consuming. Concern arises when the misleading or false information, called "disinformation", is exposed to many online platform users who have potential to react on it. The spread of disinformation can cause malicious consequences such as damage to critical infrastructure networks such as electric power, gas, and water distribution networks. Imagine a fake electricity discount, shared by disinformation campaigns, is exposed to many users on Twitter encouraging them to shift their electricity usage to a specific peak hour. If the population of users who engage with the fake discount exceeds a threshold, a blackout can happen due to the overconsumption of electricity. Thus, users access and exposure to accurate information on time can reinforce the infrastructures which are backbone of well-being for societies and economic growth.

In this dissertation, we propose solutions to protect infrastructure networks from disinformation campaigns. The solutions include: (i) an integrated epidemiological-optimization (EPO) model involving a mixed integer linear programming model (MIP) and SIR (Susceptible, Infected, Recovered) model to protect physical infrastructure networks by counter disinformation (accurate information) spread in information networks, (ii) a disinformation interdiction model to influence physical infrastructure commodity consumers with accurate information given the topology of social network, (ii) a robust mixed integer linear programming model to propose solutions superior to the original EPO model under uncertain spread of disinformation scenarios.

We illustrate our proposed models with two different case studies: (i) a sub-network of the western US interconnection power grid located in Los Angeles County in California, and (ii) the New York City subway system.

Chapter 1

Introduction

1.1 Integrated Information-Infrastructure Networks

Information networks helps people to communicate. It comprises interconnected communication devices such as computers, servers, and routers to transmit data and information between users. The number of people joining social media platforms such as Meta (Facebook) and YouTube are growing every year (Ortiz-Ospina and Roser, 2023). Even though social media platforms help people to connect with friends and family, promote businesses, and access data and information in real time, the false and misleading information spread on online platforms can negatively affect people daily lives. For instance, imagine a Tweet, widely spread on Twitter, offering fake discount during electricity usage peak hours might end up with massive number of electricity consumers shift majority of their electric power usage to peak hours and damage electric power infrastructures. Consequently, the physical infrastructures relying on electrical equipments might encounter disruptions too. For example, gas stations cannot pump gas into the cars, water well pumps cannot pump water into the houses, and traffic signals stop working. In short, the spread of false and misleading information in social networks at a trivial cost, can damage daily lives drastically. Therefore, preventive strategies to counter disinformation prevents the commodity consumers to engage with disinformation and damage physical infrastructure networks. In this dissertation, we

focus on exposing social network users to accurate information and make them aware of disinformation so that we protect physical infrastructure networks.

1.2 Models and Formulations

To illustrate how disinformation campaigns can impact the physical infrastructure networks, we conceptualize how malicious actors in information networks can threat the physical infrastructure systems by weaponized disinformation in Figure 1.1. This figure involves two different layers: (i) an information layer comprising social media users, shown as nodes, and their virtual interactions, shown as links, and (ii) a physical infrastructure layer involving the connected components such as electric power generators and gas storage, shown as nodes, connected to each other by links such as trasmission or distribution lines for electric power networks or transmission pipelines for gas networks. The links connecting these two layers are the user behaviors influenced by users (actors) in the information layer and manifested in the physical infrastructure layer. A weaponized disinformation campaign triggers in information layer by a set of users and spreads throughout the network by communication, say to the users shown by red nodes whose their responses to disinformation (e.g., consumption behavior shift) negatively impact the performance of the physical infrastructure layer.

Several different strategies are developed to combat disinformation such as empowering social network users to fact-check online contents (Rehm, 2018; **?**; **?**), imposing regulations to prevent the spread of disinformation by users (Santos et al., 2022; Zhu and Yang, 2023; Caled and Silva, 2022), developing fact-checking platforms by tech companies (Hasan et al., 2021; Mheidly and Fares, 2020), and exposing the users to accurate information and clarify disinformation (Jamalzadeh et al., 2022; Waniek et al.,

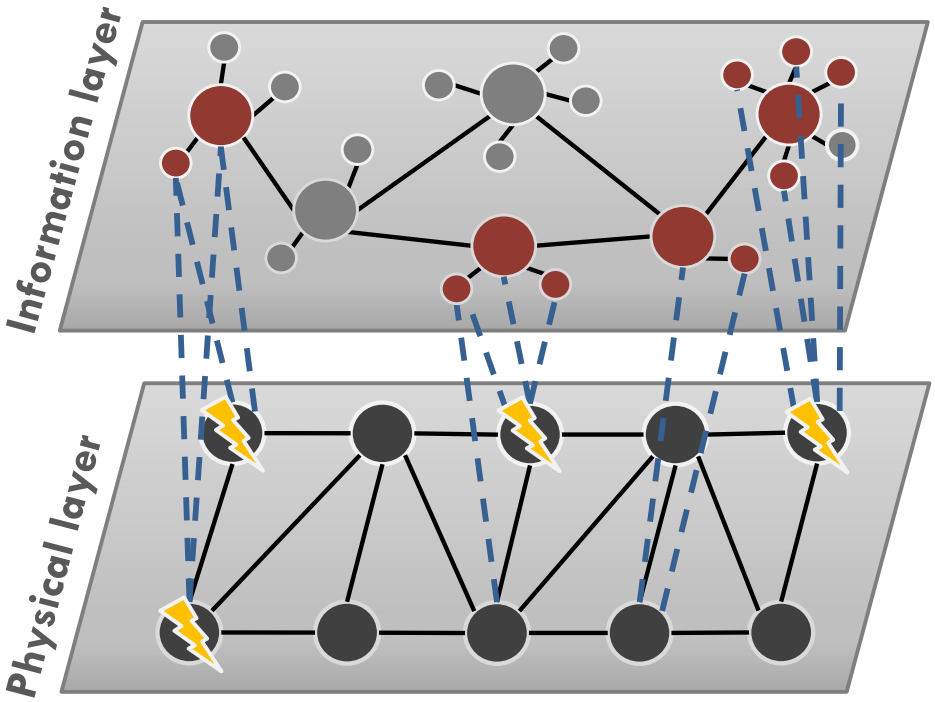


Figure 1.1: Interdependent information and physical infrastructure networks, where disinformed users shown by red nodes in the information network can disrupt nodes in the physical infrastructure network.

2021). We focus on the last strategy in this dissertation. Exposing the users to accurate information and estimating their behaviors is formulated in several different ways. *Social influence maximization model* is the first example of the kind, where the users behaviors is influenced by the other users whom they communicate with (Kumar et al., 2022; Lotf et al., 2022). *Competitive influence model* is the second category of models which deals with several different actors compete with each other to influence their target users the most (Kermani et al., 2017). The third category of models is called *information diffusion model* which obeys the Tobler's first law of geography: "everything is related to everything else, but things near are more related than distant things (Tobler, 1969)". In information diffusion context, the social network users are more likely to be convinced to not engage with disinformation by their relatively closer

connections rather than the farther ones (Carnes et al., 2007). The fourth category of models is called *compartmental* or *epidemiological* models (Kermack and McKendrick, 1927). The *Susceptible-Infected-Recovered* (SIR) model is a foundational model which is extended further to involve more than three groups of population (i.e., S, I, and R). In information context, the letters S, I, and R represent three different types of users including the users who are susceptible to receive disinformation (S), the users who are engage with disinformation, and the users who perceive disinformation (R). These models and their variants are used in different contexts and motivated researchers to predict and influence social network users and build a competitive advantage (Chen et al., 2009; Li et al., 2018).

Physical infrastructure networks are subject to risks. The risks can originate from different resources such as natural disasters and cyberspace threats. In 2012, the Hurricane Sandy damaged electric power transmission and distribution lines in New York City and caused cascading interruptions to other interdependent infrastructures such as telecommunication and fuel networks (Haraguchi and Kim, 2016; Kaufman et al., 2012). In 2020, a ransomware (i.e., a malicious malware that holds data and information hostage) caused a US natural gas compression facility to shutdown the operations for two days (ran). Such threats to physical infrastructure networks have been motivative the researchers to develop infrastructure resiliency reinforcement strategies. For example, Liao et al. (Liao et al., 2018) proposed an integer programming model to protect transportation networks under disasters with limited budgets. Najarian and Lim (Najarian and Lim, 2020) developed an integer programming model to maximize the resiliency of infrastructure networks by introducing a novel formulation to incorporate utility function into components enhancement. Then, it suggests the allocation of limited budget to enhance infrastructure components. The resiliency enhancement optimization models are also extended to interdependent infrastructure networks. Karakoc

et al. (Karakoc et al., 2023) developed a multi-objective mixed integer programming model to enhance interdependent network resiliency with a minimum budget. The solution to the model suggests pre- and post-disruption investments strategies to maximize the resiliency and minimize resource costs.

The examples of past infrastructure threats are numerous in literature and is possible to be more in the future, which is uncertain. In information and physical infrastructure optimization models the inputs and parameters to the models are unknown or estimated inaccurately in advance, before the models be solved. Thus, the solutions proposed by solving the deterministic infrastructure protection problems before the threat incidents might not be optimal for the time of the threat. To conquer the uncertainty, the optimization models can be solved for a range of plausible realizations of disinformation spread scenarios, using the robust optimization models (**?**Ben-Tal et al., 2009). In recent years, robust strategies are highlighted by researchers to combat misleading contents (Ali et al., 2021; Horne et al., 2019; Mahabub, 2020). Still, a robust optimization model to protect physical infrastructure networks during disinformation campaigns is not developed yet. One of the contributions of our dissertations is a mixed integer robust optimization model proposed to minimize the harmful impacts of disinformation campaigns on physical infrastructure networks.

1.3 Structure of the Dissertation

Our contributions are presented in Chapters 2, 3, and 4.

In Chapter 2, we present an mixed integer programming model to minimize the harmful impacts of disinformation campaigns on physical infrastructures. The proposed model integrates two models including an information epidemiological model and a physical infrastructure network flow optimization model. The former model is

formulated by SIR (Susceptible-Infected-Recovered) model which represents the evolution of population who engage with disinformation, and the latter model plans for targeting accurate information to combat disinformation campaigns and balance the performance of physical infrastructure networks. The proposed model is illustrated with a subset of the western US electric power grid bounded by Los Angeles County in California.

In Chapter 3, we propose a mechanism to minimize the spread of disinformation in social networks that damages physical infrastructure networks. Our proposed model is motivated by a nonlinear competitive information cascade model (Carnes et al., 2007). We reformulated the nonlinear integer programming model into a linear mixed integer programming model to be able to solve large-scale problems efficiently using commercial solvers. The proposed model is tested on the NYC subway system.

In Chapter 4, we propose a robust mixed integer programming model to protect infrastructure networks from a range of harmful realizations of disinformation campaigns. The proposed model combats disinformation with a limited budget by exposing deceived population to accurate information so that the manifested impact of disinformation campaigns on physical infrastructure networks is minimized under uncertain propagation of disinformation. The solutions to the robust optimization model supercede the solutions to the original model in Chapter 2 under uncertainty.

In Chapter 5, we present a summary of conclusions and future directions.

Chapter 2

Protecting Infrastructure Performance from Disinformation Attacks

2.1 Introduction and Motivation

Well-publicized disinformation campaigns surrounding recent US Presidential elections and the adoption of pandemic-related vaccinations have increased awareness among researchers that historical problems of misinformation / disinformation are exacerbated due to the wide availability and use of online platforms. Disinformation, defined as information that falsely characterizes the state of the system, including rumors, factual errors, and attempts at deception (Floridi, 2005), is rising on online platforms (Vosoughi et al., 2018; Allcott et al., 2019).

There is substantial literature on modeling the effects of and protection against false data injections by adversaries and connections to the operability and functionality of critical infrastructures (Huang et al., 2019; Liang et al., 2015; Raman et al., 2019). However, an over-the-horizon problem may result from an adversary that seeks to attack critical infrastructure indirectly by altering the consumption behavior of human intermediaries who are influenced by weaponized disinformation distributed by the adversary.

Consider the following plausible scenarios that are extended from collections of actual events. An airline passenger could tweet an alert about a suspicious package,

which, if shared rapidly and widely, could cause significant delays in flights and major traffic jams on many primary, secondary, and tertiary roads (similar to what was experienced at London's Gatwick airport (DLY)). Hackers could compromise a major US pipeline network, but the rampant spread of misinformation leads to a dramatic escalation in the aftermath and a physical, real-world increase in gas prices (similar to what was experienced when a news network spread a false story about a Russian hack of the US power grid (Molina and Sundar, 2019)). Finally, false reports of accidents on social media could lead to dynamic rerouting of drivers, causing congestion in particular areas subject to attack (similar events have occurred globally (BOT; JAM; Waniek et al., 2021)), and could worsen with the emergence of autonomous vehicles (DeBruhl and Tague, 2018)).

To begin to address some of these scenarios, we develop a model to examine the interactions between information/disinformation spread, subsequent commodity consumption behavior, and the resulting infrastructure network balance. To do so, we integrate an epidemiological model of information/disinformation spread with a network flow model. We relate the two models through human intermediaries who adopt information/disinformation that changes the way that they interact with the infrastructure network.

Contributions of our integrated model include: (i) we account for the evolution of disinformation spread over time based on the outcomes of virtual interaction between pair of users in social media, ultimately projecting user behavior onto commodity consumption, (ii) we introduce an information protection mechanism to combat against disinformation spread, and (iii) we develop a mixed-integer programming formulation to balance the performance of the critical infrastructure network and plan for targeting (good) information to counter disinformation.

This chapter is organized as follows. The Background and Literature section provides a methodological background on the concepts of disinformation spread and critical infrastructure network optimization along with the associated literature. The subsequent model section explains the proposed integration of epidemiological and mathematical programming models. The Case Study section illustrates the proposed model with a case study involving the power distribution network in Los Angeles County, California. Finally, the Conclusion section offers concluding remarks and future research opportunities.

2.2 Background and Literature Review

Our proposed work relies on two key areas: (i) spread of information and disinformation, and (ii) network flow models for infrastructure. In this section, we offer a review of these two areas, detailing some of the current research gaps addressed in this chapter.

2.2.1 Models of Information and Disinformation Spread

Within social networks, people exchange information to ultimately influence others, where influence is defined as an action "to induce a change in the behavior of another that is in accordance with the wishes of the influencer" (Hamill, 2006; Hamill et al., 2007). Each individual communicates with (or influences) many other peers and, similarly, individuals are influenced by numerous other peers. This influence can take on negative forms, such as information pollution, fake news, propaganda, misinformation, misinformation, disinformation, and hoaxes (Meel and Vishwakarma, 2020; Wardle et al., 2018; Santos-D'amorim and de Oliveira Miranda, 2021). Social media users can become a source of online broadcast activity that affects personal and social behavior.

Users may help speed up the transfer rate of information or disinformation and manipulate the content to match their points of view, deliberately or inadvertently, which may not be necessarily verified or verifiable. In such an online environment, if users do not pay enough attention to verified content and reliable sources of information, the information they receive may have varying levels of correctness and malicious intent.

The community of users can be classified into different categories based on how they respond to the influence of others. An analogy to the spread of influence and different categories of response is the spread of disease and different states of infection found in epidemiology literature (Kermack and McKendrick, 1927). A basic model for the spread of disease is the susceptible-infected-recovered (SIR) model, which uses a series of differential equations to describe the membership of different states at a point in time: those who are susceptible to the disease, those who are infected by it, and those who have recovered from it. An analogy can be made for those users reacting to information and disinformation. For example, for a group of power utility users who receive a fake message promoting a discount price for power usage during a specific time, those users may potentially share it with others or not, based on characteristics (e.g., personal traits) they exhibit. Using the SIR convention, individuals who adopt this disinformation and react to it directly by consuming more power can be classified as "Infected." Alternatively, users who are not influenced by this disinformation, for any reason, can be classified as "Removed." And users who have not received notification yet can be classified as "Susceptible." This classification of categories allows us to model, quantify, and predict their power usage during disinformation dissemination. There is a rich literature that formulates the phenomenon of transition between categorical labels with SIR models that employ a system of differential equations based on mean field theory or agent-based models that allow us to simulate the transmission of

disinformation among autonomous agents in a flexible microscale manner (Sahafizadeh and Ladani, 2018; Bodaghi et al., 2019; Beskow and Carley, 2019).

Social media users are not limited to categories S, I, and R. For example, some groups of users intend to spread accurate information to fight against disinformation, or those who have already received disinformation but do not reshare it with other users, or those who have received disinformation but temporarily do not share it (Wang et al., 2021b; Zhao et al., 2012a; Han et al., 2018). There may be communities on social media that spread authenticated information to counteract disinformation (Shrivastava et al., 2020; Wang et al., 2021b). Furthermore, users in each category (that is, S, I, and R) can be classified as *aware* and *unaware*, where it is assumed that unaware users can become aware users based on contact with aware individuals at a given rate. Still, it is assumed that the reverse transition will not occur. We define the terminology "awareness" as knowledge and understanding that something is happening or exists (MER). In addition to the novel categories attached to classic SIR models to represent a community, some methods are developed to avoid bias originating from discretizing the solutions of SIR models (Rui et al., 2018).

Several different derivatives of the SIR class of models have been developed to extend the various categories of adoption of influence (e.g., information and disinformation) (Hethcote, 2000). Given the link between networks and the spread of diseases (Keeling and Eames, 2005; Pastor-Satorras et al., 2015), the SIR modeling enterprise has applications in other network-related applications: the spread of ideas (Bettencourt et al., 2006; Woo and Chen, 2016; Liu et al., 2017; Chen et al., 2019) and the influence of social networks (Woo et al., 2011; Jin et al., 2013; Wang et al., 2015; He et al., 2015; Khurana and Kumar, 2018). A related idea by (Zhao et al., 2015) uses a variation of the SIR model to address the stifling of rumors. Still, it does not adequately allow for the competitive nature to describe the spread of information versus

disinformation. It is because disinformation spreads differently than information, as noted in (Zhao et al., 2020), with the former spreading faster and covering a large population on Twitter (Vosoughi et al., 2018). Social responses to disinformation will be examined by observing (i) how people evaluate information, (ii) how varying situations affect people's ability to evaluate information effectively, and (iii) how people act on information, including redistributing disinformation. In our context, S refers to individuals who have not yet been exposed to the disinformation content, I represents individuals who have heard the disinformation and changed their consumption behavior as a result, and R represents individuals who have heard the disinformation but ignored them after realizing that the information they received was not true or accurate.

2.2.2 Models of Infrastructure Flow Optimization

Complex infrastructure systems such as water, gas, transportation, and electricity are crucial for society's well-being and for promoting economic productivity. If one component of the system is affected by failure, larger spread effects can be experienced in other networks of infrastructures and networks of community members that suffer from unmet demand for goods and services (Barker et al., 2017). As such, the resilience of critical infrastructure networks attracted researchers to study the ability of systems to mitigate the magnitude and duration of the components of the out-of-service infrastructure network (Hosseini et al., 2016; Liu and Song, 2020).

Flow balance models are developed to determine how commodities are delivered from suppliers to customers, so that performance metrics such as average unsupplied demand and transportation costs are minimized, while guaranteeing that key operational constraints are observed. By the term "commodity," we broadly refer to flows of demanded entities (e.g., electric power, water, vehicles, data, goods) transmitted from

one node to another through links connecting them. Several different flow balance optimization models are proposed in the literature that are applicable to infrastructures focused on disruptive events (Cheng et al., 2019).

In the literature, there are numerous representations of network optimization of infrastructure networks. Nien-Sheng Hsu et al. (Hsu and Cheng, 2002) presented a generalized network flow model to model the long-term supply and demand of water resources. Ayoub Tahiri et al. (Tahiri et al., 2018) proposed a network flow optimization model for similar water distribution networks, minimizing the total cost of meeting the demand for water. Alexander Martin et al. (Martin et al., 2006) optimized a gas network consisting of a set of compressors and pipes that connect the valves in order to minimize the total cost of the network subject to supply-demand balance. Mapundi K. Banda et al. (Banda et al., 2006) similarly proposed a gas pipeline network optimization model that accounted for nonlinear isothermal equations. Traffic flow optimization problems have also been proposed (Csikós et al., 2017). Darayi et al. (Darayi et al., 2017) proposed a multicommodity network flow optimization model to understand the criticality of different multimodal transportation nodes and links.

Especially important to the case study addressed subsequently are network optimization problems designed for electric power networks. Vasin et al. (Vasin et al., 2020) proposed a model to optimize the flow of energy resources through a transportation network. Costa et al. (Costa et al., 2018) developed a two-stage linear programming model to reinforce power grids against attacks on transmission lines, proposing an exact algorithm to solve the model. Leuthold et al. (Leuthold et al., 2012) developed a nonlinear mixed-integer programming model to design an electricity market such that public welfare is maximized, with an application to the European electricity market. Wirtz et al. (Wirtz et al., 2021) proposed a sustainable multicommodity system design

model with the power grid attached to the system using mixed integer linear programming. Electric power networks are critical sources of energy that enable the function of other infrastructures, and developing flow balance optimization models for electric power grids has become important for scholars (Haraguchi and Kim, 2016), as have several network flow optimization representations of interdependent infrastructure networks that include electric power (González et al., 2016; Almoghathawi et al., 2021; Ghorbani-Renani et al., 2020; Ouyang, 2014).

In this chapter, we address the challenge of how to track and respond to disinformation attacks that disrupt infrastructure networks. Embedding the evolution of disinformation diffusion intensity over time attached to a flow balance optimization model of infrastructure network has two main benefits: (i) we can monitor and analyze the performance of infrastructure network disrupted by disinformation attacks over time, and (ii) act in opposition to disinformation propagation to mitigate the effect of disruption on the infrastructure network performance. To the best of our knowledge, such a model has not been proposed in the literature yet. To address this gap, we propose a network flow balance optimization model integrated with disinformation diffusion model that enables us to take opposite actions using social media to handle interruptions in infrastructure networks caused by disinformation attack.

2.3 Proposed Integrated Epidemiological + Optimization (EPO) Model

We propose and integrate two models to examine the interactive relationship between disinformation dissemination and critical infrastructure network performance: (i) the SIR model and (ii) a network flow balance optimization model. The network balance optimization model is used to balance critical infrastructure systems with respect to

disinformation propagating on social networks, as the spread of disinformation on social networks affects the consumption behavior of social network users. These two networks are integrated in a multi-to-one environment from the social network to the critical infrastructure network, where communities of users are assigned to the set of infrastructure nodes.

2.3.1 SIR Model

We describe the disinformation propagation process in the type of modeling "compartmental models" in which the population of social media users is divided into exclusive compartments. In such a formulation, we assign the rates at which the population within one compartment is transferred to another. In general, we can classify users into three exclusive compartments over time: (i) S, the proportion of users who are unaware of the disinformation and would have acted on it if known, (ii) I, the proportion of users who consumed the disinformation and changed (acted on the disinformation) their commodity usage schedule, and (iii) R, the proportion of users who were exposed to the disinformation but either ignored or detected it and are not interested in sharing it. Dividing the population of each compartment, we can formulate the dynamics of the compartments by replacing the size of the population with the proportion of the population.

The rate of transfer from one state to another is expressed as derivatives of the proportion of population in terms of time, and we make some assumptions to express the terms of the model. As such, we have a system of differential equations that describe how the proportion of people changes across different states over time by frequent communication. For example, given a population size N, for an unaware user randomly communicating with other users, the probability that the unaware user meets a user who adopted disinformation can be expressed by $\frac{I}{N}$, and the rate of

contact can be described as a coefficient of the total population, βN. Assuming that meetings between unaware users, S, and users who adopted disinformation result in the unaware user adopting disinformation, the population of unaware users decreases by βSI, and the population of users who adopted disinformation increases by the same size in time slots. Through this transformation process, users who consume disinformation have the opportunity to detect or ignore disinformation at a rate γ, with the associated population γI, and are no longer classified as the group that already consumes disinformation. Thus, the population size that adds up to ignorant users is expressed as γI that are removed from users who consume and adopt disinformation.

We can formulate the SIR compartmental models by introducing more compartments such as hesitants, who can be the users who have not yet decided to adopt the disinformation or not. Although the introduction of new compartments can simulate the real-world information transfer process more accurately, tweaking the parameters of the basic SIR model results in different transformation evolution paths at a relatively lower computational cost. For this reason, the basic SIR model is sufficient to generate different evolutions of users who adopted disinformation at a reasonable computational cost. Estimating the proportion of population who adopted disinformation results in estimating the evolution of commodity demand changes over time horizon of interest in our proposed model.

The evolution of the proportion of these groups over time is modeled on the basis of the homogeneous SIR model, which is mathematically represented by the system of differential equations (a.k.a. mean-field equations) 2.1-2.3 subject to constraint 2.4.

$$\frac{dS_{i,t}}{dt} = -\beta S_{i,t} I_{i,t},\ \forall i \in V,\ \forall t \in T, \tag{2.1}$$

$$\frac{dI_{i,t}}{dt} = \beta S_{i,t} I_{i,t} - \gamma I_{i,t},\ \forall i \in V,\ \forall t \in T, \tag{2.2}$$

$$\frac{dR_{i,t}}{dt} = \gamma I_{i,t},\ \forall i \in V,\ \forall t \in T, \tag{2.3}$$

$$S_{i,t} + I_{i,t} + R_{i,t} = 1,\ \forall i \in V,\ \forall t \in T. \tag{2.4}$$

The index $i \in V$ represents the community surrounding node i, and $t \in T$ denotes time. Under the assumption of homogeneity, users are equally likely to interact with other users. Also, we assumed that no users leave their interactions (e.g., leave social media) during the time of analysis. Therefore, the sum of proportions of the three categories remains constant and equals 1.

The user status can change from one state to another over time. A susceptible (unaware) user encounters an infected (disinformed) user that is infected at a rate β, and a user can move from state I to R by detecting disinformation at a rate γ. That is, in essence, β governs the rate at which disinformation spreads, and γ governs the rate at which disinformed users recover their behavior. Our approach is motivated by interactions on social networks (Watts et al., 2021). However, since the parameters of the model, γ and β, govern the rate at which disinformation spreads, other means of social interaction (e.g., TV, radio, web forums) can be taken into account with appropriate rate parameter settings.

There are several ways to solve our system of equations such as the Euler and Runge-Kutta (RK) methods and their derivatives (Dormand and Prince, 1980). Each method has advantages and disadvantages in terms of accuracy order and computational cost. The SIR model that we have deployed in our analysis is a non-linear model that needs to be solved numerically by multi-stage algorithms to return the estimates with reasonable

accuracy. The forward Euler method is a special case of the RK method, so it solves our problem with relatively low accuracy. At the expense of computational cost, we found the RK algorithm suitable for solving our nonlinear system in terms of accuracy (Medvedeva et al., 2021; Tsitouras, 2011).

2.3.2 Network Flow Balance Optimization Model

Mathematical programming has proven to be an efficient approach to model and optimize engineered systems and processes (Luenberger, 1997; Bertsekas, 2015). Network flow balance optimization can be formulated into a mathematical programming model. There are different ways to formulate network flow optimization problems, however, some formulations are more efficient to solve in terms of complexity (Bertsekas, 1998). Mathematical programming problems are classified based on the type of decision variables, constraints, and objective functions used in the model. To reduce the computation costs of highly complex problems, there exist some reformulations, which help optimization algorithms to iterate relatively faster or converge to optimal solutions with relatively lower iterations. Among these models, linear programming models are polynomially solvable, while integer and mixed-integer programming models (e.g., the models with integer decision variables) are mostly computationally more expensive to solve (Li et al., 2020). Thus, modeling a problem in linear format is much better in terms of computational complexity. If integer variables need to be included in the model, there are reformulation techniques to convert or divide the models to smaller problems to be solved faster. We formulated the network flow balance optimization problem efficiently and as simple as possible to include a relatively low number of integer decision variables. As a result, we could solve the model iteratively in a reasonable

amount of time to compare the results of the optimization problem with respect to different values of model parameters.

We model the infrastructure network as a graph $G(V, E)$, where the set of nodes, V, represents the nodes incorporating demand, supply, and transmission nodes. The set of links, E, represents the links that connect the nodes. There is a link between the nodes if there is a transmission line to transmit the commodity. With the notation found in Table 2.1, the following is a mixed integer programming (MIP) model to protect the performance of the critical infrastructure network against disinformation dissemination.

$$\min_{x,h,e,d,I,g} \sum_{i\in V,\, t\in T} h_{it} \tag{2.5}$$

s.t.

$$\sum_{j\in V:(j,i)\in E} x_{jit} - \sum_{k\in V:(i,k)\in E} x_{ikt} + h_{it} - e_{it} + q_{it} - d_{it} = 0, \qquad \forall i \in V, \forall t \in T, \tag{2.6}$$

$$x_{ijt} \leq m_{ijt}, \qquad \forall i \in V, \forall j \in V, \forall t \in T, \tag{2.7}$$

$$d_{it} = p_{it} d^c_{it}\big\{(1 - I_{it}) + \{I_{it}[r^p_{it}(1 + \rho_{it}) + (1 - r^p_{it})]\}\big\}, \qquad \forall i \in V, \forall t \in T, \tag{2.8}$$

$$I_{i,t+\bar{t}_t} = I_{it} + \dot{I}_{i,t+\bar{t}_t}(1 - r_{it}\ g_{it}), \qquad \forall i \in V \setminus \{|\ V\ |\}, \forall t \in T, \tag{2.9}$$

$$\sum_{i\in V} g_{it} \leq n^p_t, \qquad \forall t \in T, \tag{2.10}$$

$$x_{ijt}, h_{it}, e_{it} \in \mathbb{R}_{\geq 0}, \qquad d_{it}, I_{it} \in \mathbb{R}_{\geq 0}, \qquad g_{it} \in \{0, 1\}. \tag{2.11}$$

Eq. (2.5) is the objective function that minimizes the total amount of commodity shortage resulting from altered consumption behavior over time. Constraint (4.47) guarantees the balance of the input, output, produced and consumed of the commodity for all nodes. The balance equations are implicitly borrowed from the model proposed

by Tang et al. (Tang et al., 2019a). Constraint (4.48) limits the capacity of the links. Constraint (4.40) represents the baseline and responsive demand in terms of the number of users targeted for disinformation given the elasticity of the commodity demand with respect to exogenous factors (e.g., discount price message). Constraint (4.50) is used to account for the counter- spread of good information as a strategy to control disinformation dissemination. Constraint (4.54) limits the number of nodes to focus information countering strategies. The last set of constraints (4.45) describes the nature of the decision variables.

Solutions to this optimization problem can guide decisions to mitigate an commodity shortage based on disinformation, namely: (i) the amount of commodity flow that should be transmitted through the links, (ii) the optimal shortage or excess in each node, and (iii) the optimal number and location of our communities (surrounding particular nodes) to spread counter information to prevent the adverse effects of disinformation campaigns.

Note that a node cannot experience a shortage and excess simultaneously at the node level. We assume that social media users react to disinformation logically. For example, once a false price discount disinformation is broadcast, social network users consume more commodity relative to their baseline usage.

Table 2.1: Model notation.

Notation	Description
Sets	
V	Set of infrastructure network nodes
E	Set of infrastructure network links
T	Set of periods
Parameters	
$\bar{t}_t$	Duration of each period starting from time t to the beginning of its next period
q_{it}	The amount of supply in node $i \in V$ at time $t \in T$
m_{ijt}	Capacity of link from node i to node j at time $t \in T$
p_{it}	Community size surrounding the node i at time $t \in T$
d^c_{it}	Commodity consumption per capita by the community surrounding the node i at time $t \in T$
r^p_{it}	Proportion of commodity consumption of the community surrounding the node $i \in V$ at time $t \in T$ responsive to price shift
ρ_{it}	Estimated sensitivity of commodity consumption of the community surrounding the node $i \in V$ at time $t \in T$ based on the price shift
$\dot{I}_{it}$	Local derivative (change per unit of time interval $\bar{t}_t$) of the proportion of community surrounding node $i \in V$ targeted by disinformation at time $t \in T$
r_{it}	The proportion at which $\dot{I}_{i,t}$ can be changed by spreading counter (good) information for the community surrounding the node $i \in V$ at time $t \in T$
n^p_t	Total number of target locations informed by counter (good) information at time $t \in T$
Decision variables	
x_{ijt}	The amount of transmitted commodity (flow) from node $i \in V$ to node $j \in V$ at time $t \in T$
h_{it}	Shortage (undersupplied) amount of commodity at node $i \in V$ at time $t \in T$
e_{it}	Excess (oversupplied) amount of commodity at node $i \in V$ at time $t \in T$
d_{it}	Nominal demand of commodity in node $i \in V$ at time $t \in T$
I_{it}	Proportion of community surrounding node $i \in V$ at time $t \in T$ adopted disinformation
g_{it}	$=1$ if counter (good) information is released for the surrounding community in node $i \in V$ at time $t \in T$; $=0$ otherwise

2.4 Case Study: Performance of an Electric Power Network under Disinformation Attack

An electric power system is a network of electrical nodes, such as power plants, transformers, or demand points, connected by links that represent transmission lines, cables, or transformers. In such networks, the nodes represent the equipment of the power system and the links are the pathways for the transmission of electrical energies. The electrical energies that are transmitted are called power flows. Electrical power networks are used to satisfy the needs of load nodes (demand nodes) anywhere in the network by transferring the electric power produced by generators (supply nodes) through the links in the network. Each link can carry the maximum commodity through the network, called the flow capacity. Therefore, flow transmissions are limited due to flow capacities in the network. Since the flow capacities are finite in power systems, the problem of transferring the flows to satisfy the demand nodes is a vital network optimization problem that needs to be studied.

To balance the electric power system, several different models and methods are proposed in the literature, such as mathematical optimization and machine learning (Fang et al., 2011). For example, Nasrolahpour et al. (Nasrolahpour et al., 2012) developed a mixed-integer programming model to alleviate electric power congestion in transmission lines to ultimately minimize the electric power shortage and total cost. Clack et al. (Clack et al., 2015) developed a linear programming model for electric power balance given its engineering requirements. In the models mentioned above, the common constraint of concern for the authors was a system-wide constraint to guarantee the balance between supply and demand nodes over the network. Also, to obtain realistic solutions to the model, the capacity of the transmission line is specified before optimizing the model. We utilize similar constraints from the literature and

include a mechanism to counter disinformation dissemination to defend the spread of disinformation.

In recent years, a handful of papers have begun to address the potential for disinformation to affect commodity consumption. Nguyen et al. (Nguyen et al., 2019) developed a vulnerability assessment model to mitigate the adverse effects of disinformation on load shedding. Tang et al. (Tang et al., 2019a) developed an optimization model to minimize total load shedding in a power network under the condition that users react to price disinformation, relating those reactions to user personality traits. Raman et al. (Raman et al., 2019) developed an attacker-defender optimization model to mitigate strategic urban power distribution system attacks based on price disinformation (e.g., falsely offering prizes for rescheduled power usage) propagated through the community based on the "Believe, Accept, and Follow Through" mechanism.

Among all critical infrastructures, the electric power grid has been attractive to scholars for several reasons: (1) the electric power grid has been at great risk of attack and threats tremendously (Tian et al., 2020); (2) electric power grid has been relatively more expensive than other infrastructures (Kovendan and Sridharan, 2017); (3) the electric grid is one of the most vital infrastructure during disasters (e.g., Hurricane Sandy in New York) since it is an indirect critical source of commodity for other vital sectors (Haraguchi and Kim, 2016). For these reasons, the analysis of the performance of power grids under dissemination of disinformation has attracted the most attention.

Although we offer a general modeling approach that can be manipulated for a variety of critical infrastructures, our case study is motivated by the electric power grid. Disruptions to power distribution systems can result in substantial economic and social costs (Garcia Tapia et al., 2019). Due to various social factors (e.g., human mistakes, irrationality, intentional gaming, malicious attacks), the electric power grid may become more vulnerable to various kinds of cyber and physical activities when social

information becomes tightly integrated into its operation. For example, a coordinated attack could cause a significant impact, such as that experienced in the Ukrainian power grid in 2015 (UKR).

Electric power utilities are increasingly taking advantage of *demand response* programs to reduce or shift electricity usage during peak periods in response to time-based rates or other forms of financial incentives to customers (Tang et al., 2019a; Lund et al., 2015). Such demand response programs will be important in the future grid (Schuitema et al., 2017; Raman et al., 2020). Demand response messaging has been primarily textual, coming from text messages, emails, or other social media messages (Raman et al., 2019; Jain et al., 2015). Naturally, these messages affect human consumption behavior and are used to run an efficient electrical power grid system. Unfortunately, this valuable and effective mechanism could also be used to spread *disinformation*, thus creating a weapon to create a harmful effect - disrupting the power system. Imagine a Tweet being spread by a realistic but fake Twitter account. A discount price is offered to those whose power usage exceeds their average daily use by 30% during summer afternoons. As more and more customers (even those who are not the creators of the disinformation) spread this disinformation and subsequently adopt its message, blackouts will occur more likely due to overloads in the system, along with broader spread impacts to public health and safety. If a threshold of users in a particular geographical location adopts disinformation, a disruption in the power network will occur.

2.4.1 Determining the Parameters of the Model

The proportion of social media users who may adopt disinformation is found in the SIR model as the proportion of users that make up the group S.

To evaluate the relationship between this spread of disinformation and the demand for electricity, we must also estimate the change in the demand for electricity driven by

users whose consumption changed based on disinformation. The elasticity of the use of electric power measures the responsiveness of the electric power demanded to a change in price. Data from the US state level show that the estimated residential price elasticity of electric power demand is –0.7, suggesting that residential electricity consumption is inelastic to price changes (Miller and Alberini, 2016; Burke and Abayasekara, 2018). A well-known formula, the midpoint method, used to compute the elasticity of electric power demand is found in Eq. 2.12, where $pr_{i,t}$ represents the price of the electric power utility at node $i \in V$ at time $t \in T$. As a result, the parameter $\rho_{i,t}$ estimates the proportional change in electric power usage around node $i \in V$ at time $t \in T$ based on a false discount price message.

$$\rho_{i,t} = \frac{d_{i,t+1} - d_{i,t}}{pr_{i,t+1} - pr_{i,t}} \times \frac{pr_{i,t+1} + pr_{i,t}}{d_{i,t+1} + d_{i,t}} \tag{2.12}$$

Disinformation messages can take different forms to affect electricity use, such as fake weather conditions, fake availability, and price of alternative sources of electricity (e.g., coal, oil, renewable sources), and false announcements that describe the general economic situation. Such factors have been shown to significantly influence electric power consumption (Borenstein, 2009; Wang et al., 2021a) with different effects depending on geography and spatiotemporal aggregation (Miller and Alberini, 2016; Burke and Abayasekara, 2018).

The management of loads in electric power systems is highly dependent on the retail price and sales of electricity. Broadcasting false price discount signals on social networks results in increased electricity use by a large number of consumers at once, which may eventually lead to overload or blackout at the node or system level. However, when users who choose to adopt the false pricing message receive it, not all electrical power usage changes accordingly. That is, not all electric power usage is responsive to

price changes. For example, price changes might affect a more variable usage appliance (e.g., air conditioner usage may increase) relative to one that is not as variable (e.g., a refrigerator will use the same amount of electricity regardless). Based on the US Energy Information Administration, we assume that 92% of electricity use is responsive to price changes and the rest is consumed to meet basic needs of life (EIA). Therefore, we assume that a false discount price message can only affect 92% of electricity usage.

Additionally, not everyone has the ability to receive disinformation messages on social networks. The reports reveal that 82% of the US population were social media users in 2021 (STT). Therefore, we assumed that in each population surrounding electric power buses, only 82% people have access to social media directly and are susceptible to receiving the disinformation message. Among these communities, not all residents are sensitive to disinformation about the change in electric price, as not all family members are responsible for household decisions. It is shown that the composition of the home is an influential factor in determining electricity consumption (Brounen et al., 2012). Thus, we adjust the community p_{it} accordingly to incorporate the effective proportion of the population who may be more responsible for household decisions and who may be more at risk of adopting disinformation.

In summary, we assumed that a fake discount price message can affect 82% of users, and among the total electricity usage they consume, only 92% of their usage may change based on the disinformation they receive (if they adopt the disinformation). As a result, the demand per capita, $d^c_{i,t}$, is modified in the model accordingly.

Designers of electric power transmission systems should ensure that the system can operate normally under unanticipated loads. The safety factor is used to tolerate the system to meet unexpected loads and avoid waste of energy resources. Furthermore, the safety factor of electric power transmission systems is a measure of transmission line reliability that accounts for the possibility of overloaded electric power flow through

transmission lines. Due to the lack of data, we use a primary Linear Programming (LP) model with no specified capacity to estimate the capacity of the transmission lines. Assuming that the power distribution lines operate optimally, we take 0.2 ($\pm 20\%$) as a proportion of the allowed volatility of the power flow in the transmission lines. We set the estimated values as the capacity of the transmission lines.

The model we proposed to estimate the capacity of transmission lines is as the following LP problem in Eqs. 2.13-4.57, where $\bar{d}_{i,t}$ is the nominal demand of the population assigned to the bus i at time t.

$$\min_{x,h,e} \sum_{i \in V,\, t \in T} h_{it} \tag{2.13}$$

s.t. (2.14)

$$\sum_{j \in V:(j,i) \in E} x_{jit} - \sum_{k \in V:(i,k) \in E} x_{ikt} + h_{it} - e_{it} + q_{it} - \bar{d}_{it} = 0, \qquad \forall i \in V, \forall t \in T \tag{2.15}$$

$$x_{ijt}, h_{it}, e_{it} \in \mathbb{R}_{\geq 0}. \tag{2.16}$$

2.4.2 Numerical Results

We applied the proposed model to evaluate the effect of disinformation on electric power distribution systems. The electric power nodes supply the electricity to its surrounding community. We overlay the geospatial population data with the topology of the power network to establish the boundary of the model. We use the US Census Application Programming Interface (API) to collect geospatial population data surrounding each electrical power node (API). The American Community Survey (ACS) provides population data, for use in the model. In addition, we spatially clipped power nodes and links (that is, we generated a shapefile based on spreadsheet data provided

by the synthetic power grid data set (WIM)) that intersected with LA County block groups and overlaid it with LA County population data. Approximately 1600 people live in each block group in LA County.

For homeland security purposes, the actual topology of the power grid is not publicly available. However, the information on topology, demand and supply is estimated using *network imitating method based on learning* (NIMBLE) (Soltan et al., 2019), resulting in a publicly accessible synthetic power grid data set for the western interconnection grid of the United States (WIM). We limited the boundaries of this larger power distribution system to Los Angeles (LA) County in California. It incorporates more than 500 power buses (nodes) and 600 transmission lines (links), the geospatial dispersion of which is shown in Figure 2.1.

Although studying the effect of disinformation evolution on critical infrastructures at the micro-level makes more sense, highly detailed information is not publicly available. For example, while studying the evolution of disinformation integrated with low-voltage electric power systems may be more desired, we may be more able to estimate parameters of high-voltage electric power systems across a relatively larger geographical area (e.g., the treatment of the impacts of disinformation on a high-voltage network by (Tang et al., 2019b)).

A portion of commodity consumers are assumed to be users of social networks. We relate consumer products to spatially defined block groups, defined as a statistical division of US Census tracts that consists of clusters of blocks that generally contain 600 to 3000 residents of the contiguous area (USC), and each group of consumers is linked to the nodes explained.

To relate social media users to electric power buses, we performed geospatial operations for these two features. First, we defined two sets that incorporate electric power

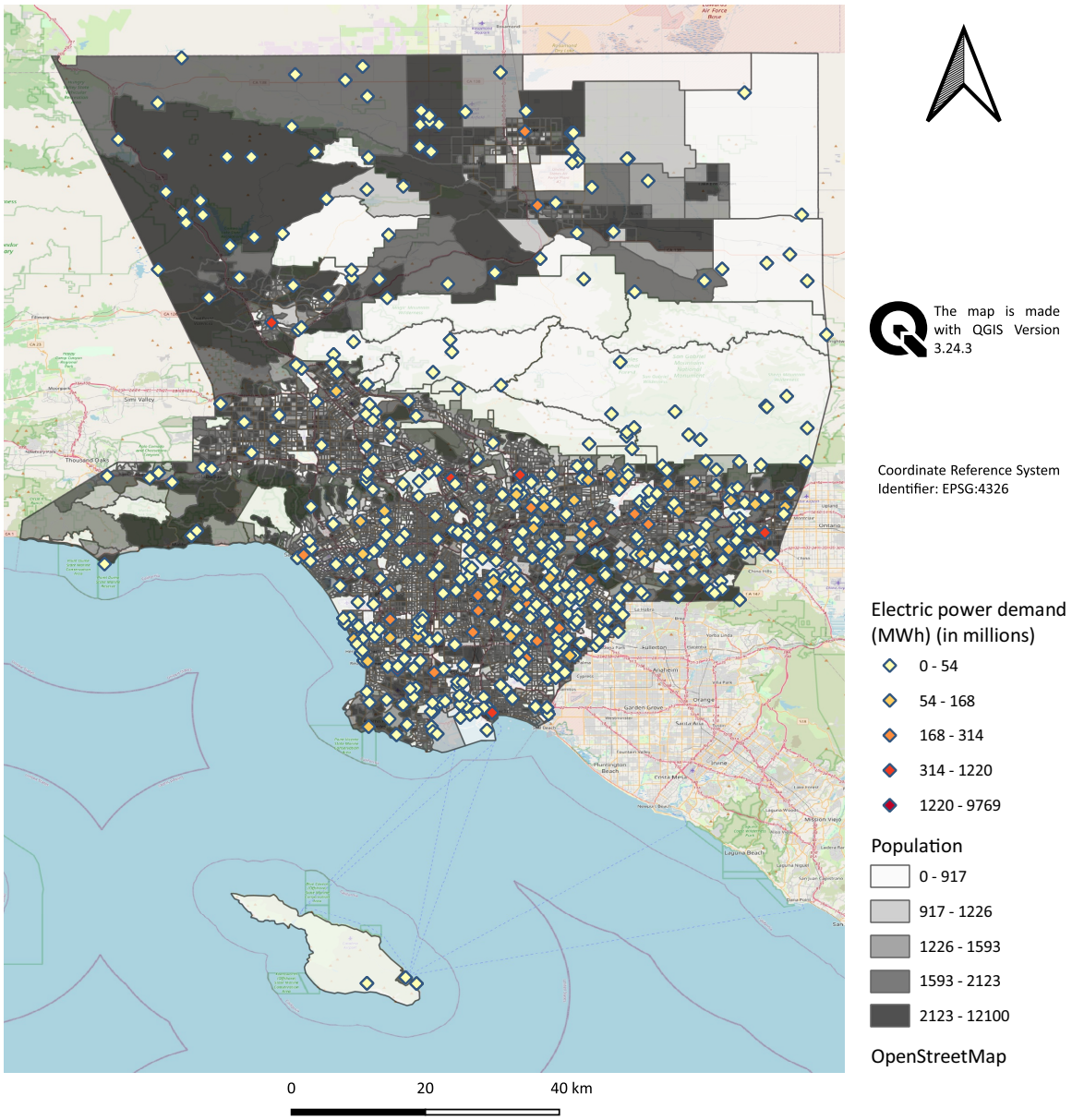


Figure 2.1: Distribution of electric power demand and population, LA County, USA.

buses and aggregated social media users. Electric power buses and social networks are geographically represented by points and polygonal features, respectively. Social media users live in census block groups, defined as a statistical division of US Census tracts that consists of clusters of blocks that generally contain 600 to 3000 residents of the contiguous area (USC). Then we calculated the Euclidean distance matrix between

the electrical power buses and the centroid of polygons. As a result, block groups are assigned to one power bus according to their shortest Euclidean distance, and several block groups are mapped to electric power buses and are characterized by estimates of power usage. As such, SIR models are deployed for each social media user within each block group.

Based on the estimated usage of social networks in 2021 (STT), it is assumed that 82% of the population in each block group have active access to social networks, so they are potentially susceptible to being targeted by disinformation. To run the disinformation propagation model, we considered 1% population being targeted by disinformation at the beginning of the analysis time period. As time goes on, the proportion of susceptible, infected (targeted) and recovered users changes according to the contact rate, the rate of being targeted by disinformation, and the rate of being aware of disinformation.

The topology of the power distribution network and the community layer are integrated as a one-to-many setting such that many block groups are assigned to one and only one power bus based on their shortest Euclidean distance to the set of power bus candidates. In other words, the population in block groups is clustered such that the locations of power buses are set as the mean of population clusters.

We interpret the parameters β and γ as the rate of disinformation degree of interest and the rate of awareness, respectively. A higher value of β results in a higher number of people targeted by disinformation per time period. The higher the value of γ, the larger the number of people who become aware of disinformation after being targeted per time period. We analyzed the sensitivity of the solutions for different values of β and γ, as shown in Figure 2.2. In these graphs, the ideal condition for β and γ is in the upper left corner of the figures, where the rate of degree of interest takes on the lowest value and the awareness rate is set to its highest value. On the other hand, the

worst case is where β is relatively higher and γ is relatively lower, which is located in the lower right corner of the figures. We ran the model with respect to several different instances of the values β and γ to analyze the sensitivity of the total shortage, the total number of communities targeted by counter information, the total flow, and the total infected communities. The results are normalized to show the percentage of difference in the resulting values.

To measure the potential spread of disinformation between social media users, the basic reproduction number ($R_i^0 = \frac{\beta}{\gamma}$) is used. It reveals the expected number of secondary susceptible users that a targeted user can affect with disinformation. For example, given $R_i^0 = 20$, each newly targeted user is expected to affect 20 secondary users within the community i, assuming that all other users contacted are susceptible. To eliminate disinformation or decrease the number of targeted users, the basic reproduction number should satisfy $R_i^0 < 1$, otherwise disinformation spreads over the network over time ($R_i^0 > 1$), or the number of targeted users remains constant over time ($R_i^0 = 1$). These concepts help to understand the situations in which the numerical results are analyzed.

We use the differential equation package (JLD) in Julia programming language to solve the SIR model. We used the mathematical optimization modeling language JuMP (Dunning et al., 2017) to codify our optimization problem in Julia with the optimizer, CPLEX (JLC), attached to it. Also, we generated the map in Figure 2.1 using QGIS Version 3.24.3 (GIS).

Based on the assumptions discussed previously, we run the model several times to evaluate decisions in different situations of spreading disinformation. The values of the parameters we used to run the model are listed in Table 2.2, and the results with different combinations of β and γ are plotted in Figure 2.2. As a validation exercise, we optimize the model for different values of disinformation adoption and

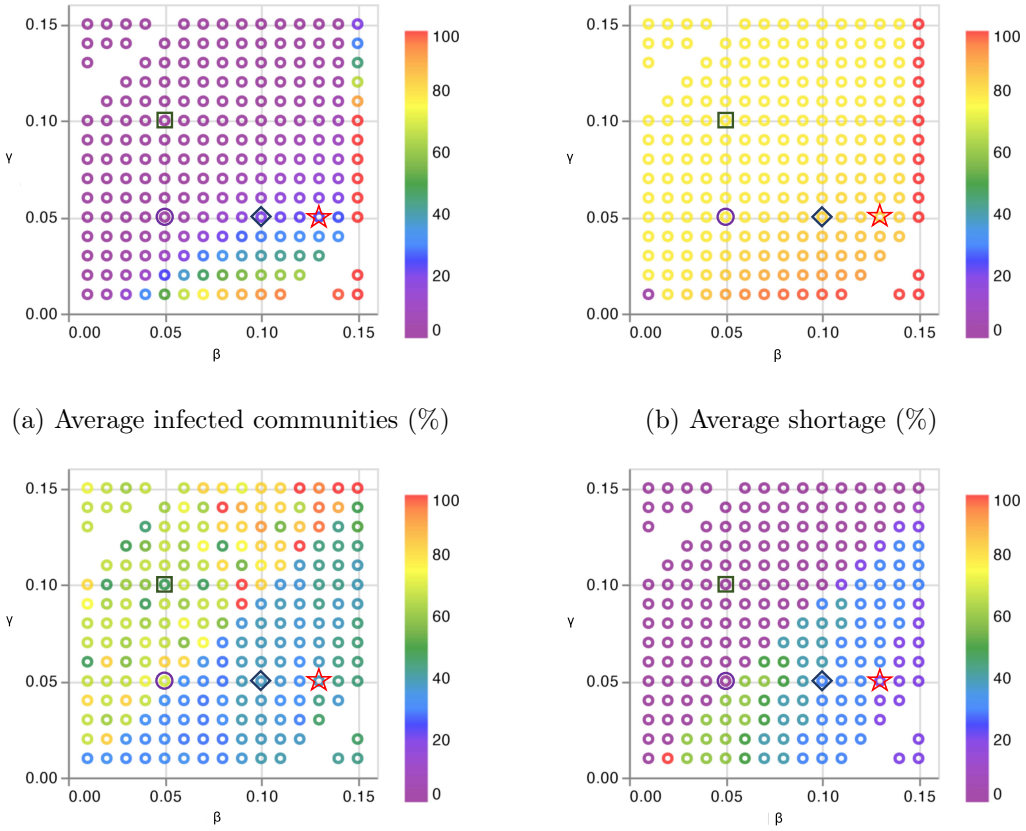
detection rate based on Raman et al. (Raman et al., 2020), who conducted a survey with more than 5000 participants to assess the proportion of people who are expected to adopt and spread disinformation about electricity prices through social networks. They evaluated different scenarios and mapping functions (i.e., linear, quadratic, cubic) to simulate disinformation spread. For simplicity, we adopted the midpoints of the simulated follow-through rates across the mapping functions and performed sensitivity analysis to illustrate the response of the metrics to different scenarios of disinformation propagation in the range. We sampled some notable instances of β and γ combinations (that is, marked by square, circle, diamond, and star shapes) to analyze the evolution of the corresponding metrics over time plotted in Figure 2.3.

Figure 2.2(a) represents the percentage of infected users in the network with respect to the governing rate of degree of interest in disinformation (β) and awareness (γ). Note that there are some empty spots in this figure (and the rest of the figures) as the SIR model is infeasible for some parameter values. The average number of infected users decreases as the awareness rate increases, although the degree of interest rate is sufficiently high in most regions. For the lower awareness rate, there is more potential to have infected communities, and it increases further for higher degree of disinformation of interest.

Figure 2.2 (b) shows the average network shortage with respect to the governing rate of degree of interest and awareness of disinformation. The average shortage increases as the degree of disinformation interest rate increases for a fixed awareness rate. On the other hand, for a higher awareness rate, the average shortage is lower while the degree of disinformation of interest level is fixed. This trend makes intuitive sense, as we saw in Figure 2.2(a) that the average infected communities decrease with higher levels of awareness, and this means that demand increases less caused by disinformation

Table 2.2: Parameter values.

Parameters	$\rho_{i,t}$	$r^p_{i,t}$	$r_{i,t}$	n^p_t	$\bar{t}_t$	horizon
Values	-0.7	0.92	0.2	10	24	169



(a) Average infected communities (%)

(b) Average shortage (%)

(c) Average flow (%)

(d) Average counter information targets (%)

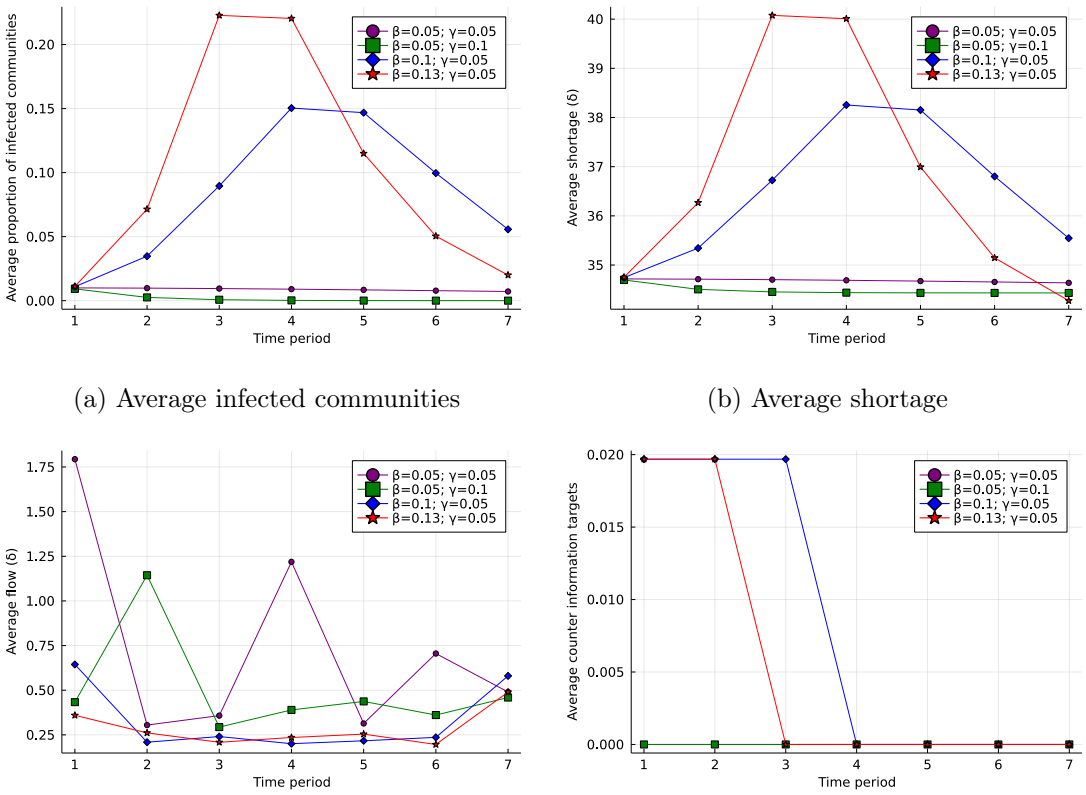
Figure 2.2: The results of (a) average infected communities, (b) average shortage, (c) average flows, and (d) average counter information targets with different combinations of β and γ as percentage of the baseline value.

and, therefore, the average shortage is reduced. The effect of the degree of interest in disinformation is greater than the awareness rate, as it always causes a shortage through the network.

Figure 2.2 (c) represents the average flow in the network with respect to the governing rate of degree of interest and awareness of disinformation. There is a clear limit in the graph where these two rates are equal ($R_i^0 = 1$ or β=γ). On the lower rectangle of values, where $R_i^0 > 1$, the basic reproduction number is high enough to allow disinformation to spread over the network in time, and on the upper rectangle of values, where $R_i^0 < 1$, disinformation has the potential to be eliminated. For the awareness rate above this bound, the average flow is higher in the network, whereas for the region below the bound, we can see much less flow in the network. This is a natural result, as we see that, based on Figure 2.2 (b), the average shortage is lower with greater awareness in the community, and this means that the network can meet demand effectively. In other words, as the average infected communities decreases for a higher awareness rate based on Figure 2.2(a), the network has more potential to satisfy demands through actual links by transmitting flows.

Figure 2.2(d) represents the average number of targeted nodes for counter (good) information with respect to the governing rate of degree of interest and awareness of disinformation. There is a clear limit in the graph where these two rates are equal ($R_i^0 = 1$ or β=γ). For the awareness rate above this bound, the average target counter information is low and not more than 20%. This suggests that if users are at least as aware as the disinformation attracts them, then we can rely less on identifying individuals with whom to supply counter information. This result is in agreement with Figures 2.2 (a) and 2.2 (b), as with the higher level of awareness, we have fewer infected communities and also less shortage, which means that less information is needed as a counter mechanism. Furthermore, there are no considerable network shortages that are problematic in this situation. For the lower awareness rate and the higher degree of interest in disinformation, where $R_i^0 > 1$, the model tries to engage more users with counter information, as shown in Figure 2.2(d).

We compared the evolution of the metrics discussed over time based on a sample combination of values of β and γ. We sampled two instances for $R^0 > 1$, one for $R^0 < 1$, and another for $R^0 = 1$. We use a normalizing constant, δ, as the units of demands and flows of electric power to interpret the output time series plots. With the time series output plots found in Figure 2.3(a) and 2.3(b), one would expect similar time series for the average infected communities and the average shortage, because an increase in the number of electric power users results in a more substantial power shortage. Figure 2.3(c) shows that despite a relatively higher peak of electric power use for $R^0 > 1$, power demands remain unsatisfied as the average flow remains low relative to scenarios $R^0 < 1$ and $R^0 = 1$. Similarly, the average flow over time remains relatively low for $R^0 > 1$, which means that the network capacity has not been used sufficiently to satisfy the demands in the corresponding scenario of disinformation propagation. This interpretation is also clear from Figure 2.3 (a), since the proportion of communities that adopted disinformation is relatively higher than in two other cases. We also observe that flows fluctuate relatively more in scenarios where $R^0 \leq 1$ is average, that is, the use of the network capacity contributed to the satisfaction of demand with a lower spread of disinformation. As Figure 2.3(d) reveals, with a higher intensity of peak demand, we do not necessarily need to target more locations to diffuse counter information. Instead, the duration of disinformation propagation plays a crucial role in selecting the number of communities that become aware of disinformation.



(a) Average infected communities

(b) Average shortage

(c) Average flow

(d) Average counter information targets

Figure 2.3: Time series of (a) average infected communities, (b) average shortage, (c) average flow, and (d) average counter information targets, with different combinations of β and γ. Note the relationship to particular points in Figure 2.2 denoted by shape.

2.5 Concluding Remarks

The proposed model aims to analyze the adverse effect of disinformation on electric power networks by integrating (i) an epidemiological SIR model to characterize the spread of disinformation in the communities surrounding electric power nodes and (ii) an electric power network optimization model focusing on minimization of the electric power shortage. In particular, we try to mitigate the effects of disinformation by identifying vulnerable power nodes and countering disinformation spread by targeting

particular communities with the spread of (good) information. To illustrate the proposed model, we solved a large-scale electric power network problem associated with Los Angeles County, California.

The evaluation of the results of our proposed model reveals how adversaries can interrupt the performance of critical infrastructures to deliver commodities to customers. In addition, we show how the intensity and duration of disinformation diffusion can be monitored to manage infrastructure performance and make communities counter the disinformation. The proposed model opens up a new space for studying the effect of disinformation diffusion in other infrastructures and managing their performance under a disinformation attack. By applying our proposed model to the large-scale electric power network, under several different scenarios of disinformation diffusion throughout Los Angeles County, we showed how our model can be applied to control the propagation of disinformation projected on the performance of infrastructures.

Due to its criticality, the electric power distribution network is used to illustrate the proposed methodology. However, the proposed integration of epidemiological and network flow models is generally applicable to a wide range of infrastructure networks with appropriate changes to the physical infrastructure flow model (e.g., physical laws that restrict the flow in gas pipelines, user behavior that affects traffic flow in a transportation network).

A primary limitation of this model is the boundary we need to draw to select an electric power distribution network to ensure a timely solution to the optimization problem. However, electric power networks are not isolated, as they interact with each other to mitigate shortages in different stations. With the evolution of computation technology, this model can be applicable and tested on larger-scale networks. Moreover, the proposed model is useful to study the effects of disinformation in other types of critical infrastructure networks, including water and gas, among others, with

appropriate physical representations governing the optimization model. Future work includes applying the proposed model to other critical infrastructure networks such as gas distribution systems, nuclear power plants, water distribution systems, etc. To extend the disinformation compartmental model, novel and flexible models (such as agent-based models) can be developed and integrated with the proposed optimization model. Moreover, some parameters used in this article are evaluated by sensitivity analysis or borrowed from previous studies or reports available online. In the future, studies will include a broader range of analysis on the fixed parameters applied to our proposed method. For example, since consumption may not vary linearly in different price ranges during disinformation spread, future work can consider the responsiveness of consumption behavior.

Chapter 3

Interdicting Disinformation to Prevent Infrastructure Disruption

3.1 Introduction and Motivation

Physical infrastructure networks are interdependent because an interruption in the operability and functionality of one network can result in cascading interruptions within the network itself and in other physically or logically connected networks. For example, an interruption in the power grid can plunge the subway into darkness and reduce performance (SUB). Physical infrastructure networks are increasingly being integrated with information technology to provide automated and efficient services to customers.

Such interdependencies need not exist only among physical infrastructure networks. A network of connected infrastructure users has interdependencies with the infrastructure itself. Information flows among the network of users can impact how they engage with physical infrastructure. For example, a tweet from a water utility warning of the need to conserve water during a particular part of the day may impact user behavior in a helpful manner. Such messages are important to new demand response programs being implemented by various utilities (Raman et al., 2019; Jain et al., 2015).

However, the spread of disinformation, or a "purposeful strategy to induce false belief, channel behavior, or damage trust" (Calo et al., 2021) on social networks, online forums, and media outlets, has led to increasingly problematic outcomes (Hunt

et al., 2022). The mitigation strategies for the COVID-19 pandemic (Larson et al., 2022; STN; Wirz et al., 2018; Cohen et al., 2022; Wong et al., 2021), the outcome of 2020 US presidential elections (Sharma et al., 2022), and the impacts of climate change (Treen et al., 2020; Fleming et al., 2021) are all recent examples wherein disinformation has swayed certain users to behave in ways leading to adverse consequences. In the aforementioned examples, the adoption of disinformation not only affected people on an individual level (i.e. loss of life due to COVID; reduced political engagement; or adverse health outcomes due to poor air quality), but also resulted in profound social consequences by eroding trust in the important pillars of modern society (e.g., medicine, science, and democracy, in general). Thus, with easily shareable disinformation, adversaries have found a new means of attack by targeting institutions and their trustworthiness.

Given how easily disinformation can be disseminated, another threat is potentially on the horizon: An adversary who seeks to attack infrastructure systems indirectly by altering the consumption behavior of human intermediaries who weaponize disinformation distributed by the adversary influences. These human intermediaries - the adopters of disinformation and users of such infrastructures - would be manipulated to interact with infrastructures in different-than-nominal ways that could lead to disruptive effects. There is a plethora of literature examining situations where such adversaries directly attack infrastructures (e.g., through viruses and ransomware (Liang et al., 2016; Huang et al., 2019)). We focus instead on indirect attacks via weaponized disinformation. Since disinformation attacks are cost-effective and easy, they have been on the rise in recent years, necessitating research that examines potential avenues for combating weaponized disinformation (CYB). A fake traffic alert could convince drivers to avoid certain locations, resulting in heavy traffic and delays in other parts of a city (Waniek et al., 2021). False pricing disinformation could lead to the overloading of the power

grid in specific locations, causing a wider spread of power disruptions (Tang et al., 2019a; Raman et al., 2020). Longer-term investments in the power grid are also at risk due to disinformation, as some media outlets blamed solar and wind energy sources for widespread power outages during the extreme cold temperatures experienced in Texas in February 2021 (TXB; TXM), when in reality renewable energy sources outperformed forecasts during 90% of the blackout (MYT). These scenarios motivate our interest in modeling mitigation efforts to interdict disinformation weaponized against infrastructure systems spread by adversaries in cyberspace.

We propose a mechanism to interdict the spread of disinformation in cyberspace to prevent unexpected human responses (i.e., changes in user consumption behavior), causing adverse impacts on infrastructure performance. This mechanism involves three main steps: (i) The interdependent structure of information and physical layers is modeled and integrated based on a one-to-many relationship so that each intermediary in cyberspace is associated with several components of the physical infrastructure affected by the intermediary, as depicted in Figure 1.1. (ii) The spread of disinformation is interdicted by counteracting disinformation broadcasting. We linearized the classic nonlinear integer programming information diffusion model (Carnes et al., 2007) and proposed a mixed-integer linear programming model, which is solvable by commercial solvers. The proposed model is estimated using branch and cut algorithm (GUR). (iii) The effect of disinformation interdiction efforts in cyberspace is projected on the physical infrastructure layer.

This chapter is organized as follows. The Background and Literature Review section provides a background on disinformation diffusion in social networks and its impact on physical infrastructure networks. The subsequent proposed Disinformation Interdiction Model section describes the proposed mathematical programming model of interdiction to minimize the harmful impacts of disinformation on physical infrastructure networks.

The Case Study section illustrates the applicability of the proposed model with a case study involving the New York City (NYC) subway network. Finally, the Concluding Remarks section summarizes the results, implications, and future research directions.

3.2 Background and Literature Review

In the present research, we explore the mechanism of how disinformation spreads in cyberspace, how to curb its reach, and how to minimize its adverse impacts on physical infrastructures. We discuss the interactions between users in the information layer and offer some formulations to quantify the outcome of these interactions and techniques to combat disinformation to protect physical infrastructures. Our proposed method focuses on interdicting the flow of disinformation based on the information layer's structure and users' availability to prevent the unexpected usage of infrastructure networks. This section reviews related papers and identifies the research gaps that guide our approach.

Physical infrastructure systems are typically modeled as networks or graphs that involve nodes (or vertices) and links (or edges) that connect the nodes. The nodes represent the assets, and the links represent the connections that enable the flow of commodities (enabling services) between each pair of nodes. Connections between components (i.e., nodes and links) from one layer to another have been studied in the literature in the form of interdependent networks (Rocco et al., 2018; Ghorbani-Renani et al., 2020; Almoghathawi et al., 2019). The connections determine how the functionality of a component in one network can affect the functionality of another network.

Recent literature has shown the vulnerability of physical infrastructure systems to disinformation propagated by human intermediaries in different contexts (Akella et al.,

2010; Peng et al., 2020; Goh et al., 2017; Huang et al., 2019). Harmful disinformation diffusion spread in social networks can lead to economic damage and productivity losses (Garcia Tapia et al., 2019). Such weaponized disinformation attacks can result in blackouts (Raman et al., 2020), heavy traffic in certain areas (Waniek et al., 2021), and interruption in political processes (Schreiber et al., 2021), among others. In these situations, the ability of physical infrastructure networks to mitigate the magnitude and duration of the failure and return to normal operation is important (Liu and Song, 2020; Hosseini et al., 2016; Mishra et al., 2016). The damages to physical infrastructure can result in cascading failures in other infrastructures with interdependent connections (González et al., 2016; Barker et al., 2017; Li et al., 2017; Zhang et al., 2021). For example, a disruption in electric power can cause disruptions in related infrastructure networks such as water pumps, traffic lights, and communication towers (Almoghathawi et al., 2021; Ghorbani-Renani et al., 2020; Lobban et al., 2021). Other examples include the injection of false data to interrupt operators' control efforts in electric power networks (Ashrafuzzaman et al., 2018; Liang et al., 2016), false pricing attacks that lead to shifts in customer power usage with adverse effects (Tang et al., 2019a; Raman et al., 2019, 2020), and the spread of false traffic information that leads to urban traffic congestion at a city scale (Waniek et al., 2021). Although the activity of human intermediaries in cyberspace is not easily observed, the propagation of disinformation in cyberspace can result in a dramatic changes in consumption behavior and subsequent damage to physical infrastructures (Mishra et al., 2016; Baidya et al., 2019; Fawzi et al., 2014). Thus, interdicting such diffusion of false information in the information layer can mitigate the adverse impact of abnormal consumption behavior in the physical layer.

Online user-generated content and its consumption are growing on the internet (NET; TWQ; SNP; INS). Such content broadcast over the web is available to all users

and can influence how they learn, think, socialize, and make decisions. Unfortunately, not only is a significant portion of users is unable to distinguish factual from false online content, but also a majority of them (i.e., 3 out of 4 users) mistakenly overestimate their ability to detect fake news shared on social network platforms (Lyons et al., 2021; USN). Given how quickly and broadly disinformation can spread, especially when contents are not verified (or even unverifiable) (TW1; TW2), user engagement with physical infrastructure networks as a result of disinformation is of concern. In such cases, dramatic disinformation spread can result in sizable consequences such as a city-wide blackout (Raman et al., 2020) or heavy urban traffic load (Waniek et al., 2021).

Tech companies employ artificial intelligence to detect disinformation and combat its spread (TWT; FBK; BRK), albeit with limited success (AI1; AI2). Other strategies might be more effective at interdicting the spread of disinformation to protect the accessibility of social network users to verified information to alleviate adverse consequences of weaponized disinformation attacks. We discuss these alternate strategies in more detail below.

There are three common approaches to modeling the diffusion of information on social networks: (i) cascade models (Kempe et al., 2003), (ii) linear threshold models (Granovetter, 1978), and (iii) epidemiological models (Rodrigues, 2016). The first two models are based on the probability that users may be exposed to and adopt disinformation, and the third is based on the influence of users relatively close to each other in a network. These approaches motivated researchers to examine how to maximize the influence of information on social networks by targeting specific users to gain a competitive advantage over rivals (competitors) (Chen et al., 2009; Li et al., 2018). Several mathematical models have been developed to this end. For example, Bharathi et al.(Bharathi et al., 2007) developed a game theory approach based on a cascade

model to influence target users by word-of-mouth propagation. Bozorgi et al. (Bozorgi et al., 2017) proposed a linear threshold model to maximize influence to target users with the minimum number of influencers (i.e., social network users with a relatively high level of trust by other users). Qiang et al. (Qiang et al., 2019) developed a mixed-integer programming model based on the linear threshold information diffusion model to estimate diffusion influence weights between users in large-scale social networks. Qiang et al.'s model predicts how social network users get exposed to information online. Epidemiological models, originally developed to represent the spread of diseases, have also been used to model the spread of disinformation. Specifically, the Susceptible-Infected-Recovered (SIR) model and its derivatives are examples of such models wherein individuals are classified into different categories: users who adopt disinformation and react to it (infected), users who are not affected by disinformation (recovered), and users who have not yet been exposed to disinformation (susceptible) (Tang et al., 2022; Wang et al., 2021c). In SIR models, coefficients govern the speed at which disinformation is propagated and detected. The literature that examines information diffusion processes allows us to develop mathematical formulations and deploy them in an optimization model to combat the spread of disinformation in social networks. As a result, we propose solutions to prevent unexpected consequences of disinformation spread related to physical infrastructure networks.

What distinguishes our article from existing work on infrastructure protection is that we represent an interdependent relationship between the information and physical layers (see Figure 1.1). We attempt to interdict the spread of disinformation in the information layer to reduce the impact of disruption in the physical layer. In other words, we propose a mechanism to stop disinformation in social networks that results in unexpected consumption behavior by users of physical infrastructure. We provide a

general formulation to address this problem and apply the model to an urban subway network to prevent maladaptive consumption behavior motivated by disinformation.

3.3 Proposed Disinformation Interdiction Model

We propose a mixed integer linear programming model to stabilize the components of the physical infrastructure network and its performance by targeting cyberspace actors that spread disinformation in the information layer.

To describe widespread disinformation diffusion in social networks, we assume that the users are more likely to adopt information (including disinformation and misinformation) received from relatively closer connections (e.g., friends) than ones with a relatively farther interaction (e.g., multiple connections away). This assumption complies with Tobler's first law of geography (Tobler, 1969), which states that "everything is related to everything else, but things near are more related than distant things." We can apply Tobler's first law of geography to the social media context, predicting that users are more likely to be influenced by the point of view of their immediate contacts than their indirect contacts (Borgatti and Foster, 2003). Support for the law is inferred when a closer connection in an individual's network exerts greater persuasive influence relative to more remote connections (Carnes et al., 2007). This law applies to our model because it targets specific users who do not want to receive disinformation from their neighboring connections.

We model the information layer (here, a social network) as a directed graph $G(V^t, E^t)$ where V^t and E^t represent the sets of users (i.e., nodes) and interactions (i.e., links), respectively. We denote the shortest interaction distance (i.e., the shortest path) between pair of social network users as $i \in V^t$ and $j \in V^t$ as $\bar{d}_{ij} \geq 0$, where $\bar{d}_{ij} \to \infty$, if user i has no interaction with another user j at all. Given the interaction distance

threshold k, we extend variable $\bar{d}_{ij}$ into a binary variable d_{ijk}, where $d_{ijk} = 1$ if user i is a kth distant neighbor follower of user j at distance k, and $d_{ijk} = 0$ otherwise. We employ this parameter to calculate the number of users adopting either accurate information or disinformation in the neighborhood of a social network user. Let A, B, and C denote users of different types as described in Table 3.1, depending on whether each social network user adopts neither accurate information nor disinformation, adopts disinformation, or adopts accurate information, respectively. We partition the set A into sets A_m and A_n to denote if the social network user who adopts accurate information or disinformation is a consumer of the infrastructure under consideration or not, respectively. The remaining consumers can be classified as members of the set B or C, depending on whether they adopt disinformation. These sets and distance parameters are used in our proposed mixed-integer linear programming model to interdict the influence of different social network users on the spread of disinformation by targeting (i.e., not receiving disinformation) a limited number of existing user accounts. Consistent with the model of Carnes et al. (Carnes et al., 2007), we assume that the users who consumed and shared the disinformation do not recover from engaging with the disinformation, even if exposed to accurate information. Thus, the intersection set of the sets A and B, specifying the infrastructure commodity users who acted on disinformation, is not considered as target users in the optimization model. This approach also helps the optimization algorithm run relatively more efficiently.

Each interaction between pairs of individual users is considered to be a link in the information layer, determining whether communication between pairs of users is established to transfer accurate information or disinformation messages. Thus, nodes and links in the information layer are associated with probabilities determining the chance that users are online and receptive to the message and the chance that an interaction tries to influence immediate connections. From this point forward, the

links used for influencing immediate connections of a particular user are referred to as an "activated interaction." The set of activated interactions shapes the structure of the information layer.

In addition to the structure of social networks, factors such as the personality traits of users, the beliefs of the recipients of information, the perceived quality and reliability of the information, the attractiveness of the message, the quality of interaction between sender and recipient, and the consistency between message and recipient beliefs affect users' decisions to be influenced by disinformation (Buchanan, 2020; Burbach et al., 2019; Leng et al., 2020; Indu and Thampi, 2020; Wolverton and Stevens, 2019; Lai et al., 2020; Tang et al., 2019a).

In social networks, personality traits can be inferred from digital footprints collected from user profiles such as demographic information and visited locations (Lambiotte and Kosinski, 2014; Azucar et al., 2018), and have been shown to affect a person's willingness to detect and discount disinformation or adopt and engage with it. We consider the personality trait model, *the big five*, that classifies traits into five categories and relates personality traits to behavior, including information behaviors. For example, neuroticism and extraversion have been shown to affect beliefs in online rumors (Lai et al., 2020). As Lai et al.'s survey results of over 11,500 social media users indicate, individuals with high neuroticism and extraversion were most susceptible to rumors. (Lai et al., 2020).

The attitudinal consistency between disinformation senders and recipients is likewise important. For example, some online users may trust disinformation shared by their contacts simply because of their shared beliefs, even though the message is not verified or verifiable (Staiano et al., 2012; Buchanan, 2020). Conversely, the information from the users with opposing attitudinal positions is likely to be discounted merely because it comes from an attitudinally discrepant source. Other factors may

also play a role. For example, the motivation of the sharer (e.g., interacting with people to feel influential) and the amount of information exchanged (Buchanan, 2020; Chen, 2016; Super et al., 2016) can also guide information sharing behaviors. We explain the likelihood of user i sharing disinformation to their immediate neighbor j in mathematical form with Eq. 3.1, based on the influence optimization problem by Kempe et al. (Kempe et al., 2003).

$$p(i \in B \mid j : (i,j) \in E^t) = f(\pi_i, \psi_{ij}, \theta_j, \zeta_i, \zeta_j) \tag{3.1}$$

Eq. 3.1 addresses three main specifications: (i) π_i describes the likelihood that user i believes disinformation, (ii) ψ_{ij} indicates the penetration (i.e., likelihood of successful transfer) level of the message between pair of users i and j, and (iii) θ_j represents the likelihood that user j shares disinformation to their immediate neighbors. Also, ζ_i and ζ_j are probabilities the users i and j are actively online and using the social network platform or are involved in the community to interact with other members. The personality traits of the user can influence specifications (i) and (iii) (Burbach et al., 2019; Lai et al., 2020), and specification (ii) can be affected by the characteristics of the disinformation message (e.g., the credibility of the source) (Buchanan, 2020). Eq. 3.1 can be generalized based on the characteristics of senders, recipients, and their interaction. For example, with a large number of online users willing to share disinformation (high θ_j) and a high likelihood of adoption of disinformation among them (high ψ_{ij}), a relatively high percentage of users will be exposed to disinformation (Qi et al., 2018). These parameters will determine the likelihood of spreading disinformation and suggest which users should be protected (i.e., blocked from receiving disinformation) to prevent disinformation dissemination among users.

Using Table 3.1, we develop a fractional programming problem in Eqs. 3.2-3.4 for identifying the optimal set of users, which are potential disinformation targets that must be protected from receiving disinformation. Then, we reformulate nonlinearities as a mixed-integer linear programming problem in Eqs. 3.5-3.11 to find solutions to the model. Solving this problem helps combat the spread of disinformation, keep the utilization of physical infrastructure stable, and deliver services with acceptable quality to consumers.

Table 3.1: Optimization model notation.

Notation	Description
Sets	
A	Users who are unaware of disinformation but would have acted on it if known
A_m	Subset of users who use the infrastructure
A_n	Subset of users who do not use the infrastructure
B	Users who consumed and shared the disinformation
C	Users who detected disinformation and shared correct information messages
Parameters	
n	Total number of users targeted to share accurate information
π_i	Personality trait score of user i
d_{ijk}	$= 1$, if user i receives information from user j (i.e., i follows j) in the interaction distance threshold k; $= 0$, otherwise
Decision variables	
x_j	$= 1$, if user j is targeted to share accurate information from their immediate neighbors, $= 0$ otherwise

$$\min_{x} \sum_{k \in K} \sum_{i \in A_m} \frac{\pi_i \sum_{j \in B} d_{ijk}}{\sum_{j \in A} d_{ijk} x_j + \sum_{j \in B} d_{ijk} + \sum_{j \in C} d_{ijk}} \tag{3.2}$$

s.t.

$$\sum_{j \in A} x_j \leq n \tag{3.3}$$

$$x_j \in \{0, 1\} \tag{3.4}$$

Eq. 3.2 is the objective function that minimizes the overall probability that users of type A_m adopt disinformation. The constraint 3.3 limits the number of users blocked from receiving disinformation. Constraint 3.4 describes the nature of the decision variables used in the model.

Proposition 1. *The fractional programming problem 3.2-3.4 is equivalent to the mixed-integer linear programming problem 3.5-3.11.*

$$\min_{x,y,z} \sum_{i \in A_m, k \in K} \pi_i y_{ik} \tag{3.5}$$

$s.t.$

$$\sum_{j \in A} d_{ijk} z_{ijk} + (y_{ik} - 1) \sum_{j \in B} d_{ijk} + y_{ik} \sum_{j \in C} d_{ijk} \geq 0, \ \forall i \in A_m, \ \forall k \in K \tag{3.6}$$

$$\sum_{j \in A} x_j \leq n \tag{3.7}$$

$$y_{ik} - z_{ijk} \geq 0, \ \forall i \in A_m, \ j \in A, \ k \in K \tag{3.8}$$

$$x_j - z_{ijk} \geq 0, \ \forall i \in A_m, \ j \in A, \ k \in K \tag{3.9}$$

$$z_{ijk} - y_{ik} - x_j + 1 \geq 0, \ \forall i \in A_m, \ j \in A \tag{3.10}$$

$$x_j \in \{0, 1\}, \quad y_{ik} \in [0, 1], \quad z_{ijk} \in [0, 1] \tag{3.11}$$

Proof. With the change of variable as Eq. 3.12, the objective function 3.2 is converted to 3.13 with a set of constraints added as Eq. 3.14.

$$y_{ik} = \frac{\sum_{j \in B} d_{ijk}}{\sum_{j \in A} d_{ijk} x_j + \sum_{j \in B} d_{ijk} + \sum_{j \in C} d_{ijk}}, \quad \forall i \in A_m, \ \forall k \in K \tag{3.12}$$

$$\min_{x,y} \sum_{i \in A_m, k \in K} \pi_i y_{ik} \tag{3.13}$$

s.t.

$$\frac{\sum_{j \in B} d_{ijk}}{\sum_{j \in A} d_{ijk} x_j + \sum_{j \in B} d_{ijk} + \sum_{j \in C} d_{ijk}} \leq y_{ik},$$

$$\forall i \in A_m,\ \forall k \in K \tag{3.14}$$

We can re-write Eqs. 3.13 and 3.14 with Eqs. 3.15 and 3.16.

$$\min_{x,y} \sum_{i \in A_m, k \in K} \pi_i y_{ik} \tag{3.15}$$

s.t.

$$y_{ik} \sum_{j \in A} d_{ijk} x_j + (y_{ik} - 1) \sum_{j \in B} d_{ijk} + y_{ik} \sum_{j \in C} d_{ijk} \geq 0,$$

$$\forall i \in A_m,\ \forall k \in K \tag{3.16}$$

Using McCormick relaxation (McCormick, 1976), we define new variables as Eq. 3.17 and convert the nonlinear integer programming model 3.2-3.4 into a mixed-integer linear programming model 3.5-3.11.

$$z_{ijk} = y_{ik} x_j,\ \forall i \in A_m,\ j \in A,\ k \in K \tag{3.17}$$

□

The ability of urban transportation infrastructure to maintain service levels is vital to community productivity (Li et al., 2019). The spread of disinformation can disrupt such service levels by manipulating user behavior. Suppose disinformation is spread on Twitter claiming a highly used New York City subway station has been closed. Subway users whose daily travel route passes through that station and who adopt this disinformation may choose an alternative route, adding more travel distance and time. When a large number of passengers adopt disinformation and change their nominal travel routes, the subway network will no longer be able to provide an as-planned optimized subway transit timetable as the demand unnecessarily shifts to other areas of the network, whereas making the passengers aware of disinformation can keep subway network performance relatively stable. To show how this interdiction helps maintain the utilization of infrastructure components, we define the physical layer and integrate it with the information layer. We denote the graph associated with the infrastructure network as $G(V^u, E^u)$, where V^u represents the set of infrastructure junctions (e.g., metro stations) and E^u represents the set of connections between junctions (e.g., metro lines). Then, we assign each user to a set of infrastructure links, E^u. Once users are exposed to disinformation that states that some infrastructure components are unusable, they start utilizing alternative components (i.e., routes) to satisfy their needs. By solving the proposed disinformation interdiction model, we show how interdiction helps with maintaining the utilization of infrastructure components (i.e., peak usage of nodes and links).

A subway system consists of a set of stations (nodes) for passengers to enter and exit the system and lines (links) that connect the stations to transfer passengers from one station to another. The subway system is used to satisfy passenger needs traveling from one origin station to the destination through the lines.

3.4 Case Study: Subway Network Performance Under a Weaponized Disinformation Attack

The impetus for this chapter is a scenario wherein a weaponized disinformation attack, attempting to manipulate the routes that passengers take to travel from origin to destination. Such an attack can consist of a list of stations that are said to be closed due to an unusual event, such as maintenance or an attack, to convince passengers to avoid those stations during their commute. The adversary disinformation spread can be deployed to either redirect the traffic of passengers from certain locations or move the passengers away from them. We refer to the former scenario as the "link-up" and the latter as the "split-up." In the first scenario, the aim is to overload some stations with passengers to reduce the quality of service of the subway system. The second scenario can be deployed to remove passengers to certain stations to overcrowd them. Under the "link-up" scenario, a station is targeted in the beginning so that if it is closed, the passengers' alternative routes pass through the most crowded station (i.e., a station with relatively higher betweenness centrality (Freeman, 1977), suggesting that such a station falls in the shortest paths of travel relatively more often). Then, the next most critical station is added to the list of closed stations, one at a time, until a predetermined number of stations is reached. Conversely, under the "split-up" scenario, a station that is part of the shortest travel paths is targeted, and the next station is added, one at a time, based on its betweenness centrality, to the set of target stations incrementally. Under these scenarios, keeping some passengers away from engaging with disinformation and thus traveling their original routes is essential, which is what we address in the model.

Social networks are instant information sharing tools for public outreach. They help transportation providers reach many passengers instantly to inform them about service

interruptions. A disinformation campaign about service interruptions can reach many passengers and manipulate their travel behavior, resulting in rerouting or leaving the subway, subsequently causing delays and subway utilization losses. Consider New York City (NYC), the most populous US city with approximately 8 million residents (POP) and known as a "tweeting town", in which residents regularly share local information on Twitter (SOC). The NYC subway–one of the largest subway systems in the world– is comprised of over 450 stations and more than 500 links (including subway lines and pedestrian connection links) with around 5.5 million daily ridership in pre-pandemic conditions (RID). With this great mass of daily ridership and large active population in social networks, a weaponized disinformation campaign could lead to extensive adverse consequences, further highlighting the need for protection against disinformation. To do this, we collected relevant NYC ridership data. Then, we applied our model to illustrate how disinformation targeting the subway infrastructure can be interdicted to protect its functionality.

By solving the mathematical model, we can target the optimal set of users to prevent them from adopting disinformation. Given that, as noted earlier, personality traits can play a role in whether people adopt (Lai et al., 2020) or reject misinformation (Wolverton and Stevens, 2019). The likelihood at which people believe in disinformation can be represented quantitatively with a rumor belief score. We simulate the scores by sampling from the multivariate normal distribution using the average and covariance matrix of rumor belief and personality traits scores collected by the survey by Lai et al. from $11,600$ participants (Lai et al., 2020). The rumor belief scores are used to estimate users who are relatively less and more likely to adopt disinformation. These scores are assumed to follow normal distribution, according to the Central Limit Theorem by satisfying four conditions: (i) The sample of participants in the survey was randomly taken online, (ii) The sample of participants was independent of each

other, where an answer of one participant in the survey does not affect the answer of another, (iii) The sample size of participants was sufficiently large (in Lai et al. (Lai et al., 2020): over $11,500$ participants), and (iv) The sample size of participants was no larger than 10% of the total population (i.e., $11,500$ out of 270 million users) and is drawn without replacement. We set the population with the highest rumor belief scores (i.e., three times standard deviation higher than the average score) as disinformation spreaders, and the lowest rumor belief scores (i.e., three times standard deviation lower than the average score) as accurate information spreaders in the social network. For the rest of population (i.e., who are unaware of either disinformation or accurate information), weights π_i in the objective function, Eq. 3.5, are set accordingly, based on the estimated likelihood of disinformation adoption (i.e., rumor belief scores).

Furthermore, personality traits may determine the likelihood that social network users like, comment, and/or share the posts on social networks (Buchanan, 2020; Burbach et al., 2019; Leng et al., 2020; Indu and Thampi, 2020; Wolverton and Stevens, 2019; Lai et al., 2020). For example, an online survey with 164 participants of different personality traits showed that almost 40% of online social network users never share online content and 37% of users never comment on any posts (Burbach et al., 2019; Buchanan and Benson, 2019); and these online behaviors have been linked to people's personality traits (e.g., "conscientious" individuals are less likely to leave comments on posts (Burbach et al., 2019)). Based on these estimations, we can assume that a user's likelihood of commenting or sharing content (with any frequency) ranges from $63 - 100\%$ (i.e., about 80% of population have potential to transfer dis/information), under complete overlap and no overlap between the population who are willing to comment and the population who are willing to share contents, respectively. We set this parameter in the model to estimate the likelihood of users' interaction willingness in advance and modify the social network accordingly.

The distance threshold, k, represents the distance between users in which, if one user shares the information, it can be viewable by another user. Since tweets and retweets can be seen on the main Twitter page for users, we set the threshold to 2. We solve the model using a single snapshot of the social network. Therefore, we only consider tweets and retweets, although disinformation can spread beyond the distance threshold of 2. In addition, isolated users were removed prior to the analyses given that they do not affect other users or are not affected by others.

Assuming the circulation of a disinformation message naming several stations that have closed unexpectedly for maintenance, the list of stations could be either where the passengers pass the most to complete their travel (i.e., split-up scenario) or where, if closed, the passengers will reroute and pass through the most traveled stations, thus increasing passengers' traffic there (i.e., link-up scenario). In each scenario, the message could state the closure of stations during the morning or afternoon peak travel. As such, we consider four scenarios and illustrate how disinformation interdiction can mitigate travel delays, the unexpected rerouting, and unmet travel completions.

3.4.1 Data and Parameters of the Proposed Model

We obtained the origin-destination (OD) passenger flow data from Blume et al. (Blume et al., 2021), who estimated the proportions of passengers traveling between different stations using a Bayesian inference method. Given these proportions (probabilities) and entry-exit count data, the expected value of passenger flows between each pair of stations was calculated. We sampled the flow of passengers based on the calculated probability to track their behavior (rerouting) while exposed to disinformation during the rush hours, 6:30 AM – 9:30 AM and 3:30 PM – 8:00 PM (Eastern Standard Time), Monday – Friday (SWY).

We consider the Twitter graph of the social network (Leskovec and Mcauley, 2012; STD), which includes over 80,000 nodes and around 1.8 million interaction links collected from public sources. We randomly selected an initial seed of Twitter users who are assumed to be online, then divided them into subway passengers (A_m) and non-subway passengers (A_n).

Although our method is motivated by the interaction among social network users, every user can be an instance of dis/information sharing because information can reach out through different means of social interaction, such as television, radio, face-to-face interaction, and online social networks, other than Twitter.

3.4.2 Numerical Results

Using the aforementioned data and proposed model, we solved the interdiction optimization problem to illustrate the impact of disinformation interdiction on subway network utilization. Solving this model helped minimize the impact of disinformation spread on unexpected travel changes through the New York City subway (see Figure 3.1).

A sample of social interaction network, borrowed from the Twitter social network, is illustrated in Figure 3.2 before and after the interdiction (i.e., solving the proposed optimization problem) under the morning split-up scenario. Each user is represented by a color based on their status. The red nodes represent users who consumed and acted upon disinformation (user type B). The blue nodes are users who detected disinformation and may share accurate information messages to counter disinformation (user type C). The gray nodes represent users who are non-subway users whose engagement in disinformation does not impact the physical infrastructure. The yellow nodes represent subway users in the social network. Moreover, the mixed gray-red and gray-blue nodes represent the non-subway passengers who share disinformation and

accurate information, respectively. And finally, the mixed yellow-red and yellow-blue nodes represent the subway users who share disinformation and accurate information, respectively.

The optimization problem finds the optimal set of combined gray and yellow nodes to share accurate information to minimize the impact of disinformation on subway infrastructure. The optimal set of users to which to share accurate information is either represented in green or mixed with green color. The green nodes impact the yellow nodes the most to make them aware of disinformation and not engage with it as much as possible. The results show that the users with a relatively higher number of connections to subway users are attractive targets to share accurate information. Also, the users who have relatively more connections to the disinformation spreading users are strong candidates to be neighbored with accurate information spreaders and mitigate the impact of disinformation spreaders of their neighbors.

Figure 3.3 represents the impact of rumored subway station closures on passenger rerouting decisions in the NYC subway network. The red dots in this figure represent the subway stations that are rumored to be closed under the split-up scenario, and the blue lines connect the origin and destination of a travel by shortest path (i.e., calculated by Dijkstra's algorithm) through the connected subway stations. Each user is assigned to a route based on the travel probability data. The passengers who do not travel through the "closed" stations (according to rumor) are not impacted by disinformation, as shown in Figure 3.3(a), even though they are engaged with disinformation. On the other hand, some other passengers who engaged with disinformation might be impacted by subway stations rumored to be closed and thus reroute, as depicted in Figure 3.3(b). The latter passengers cause the passenger traffic to redistribute to alternate stations. Passengers are assumed to take the shortest alternative route from

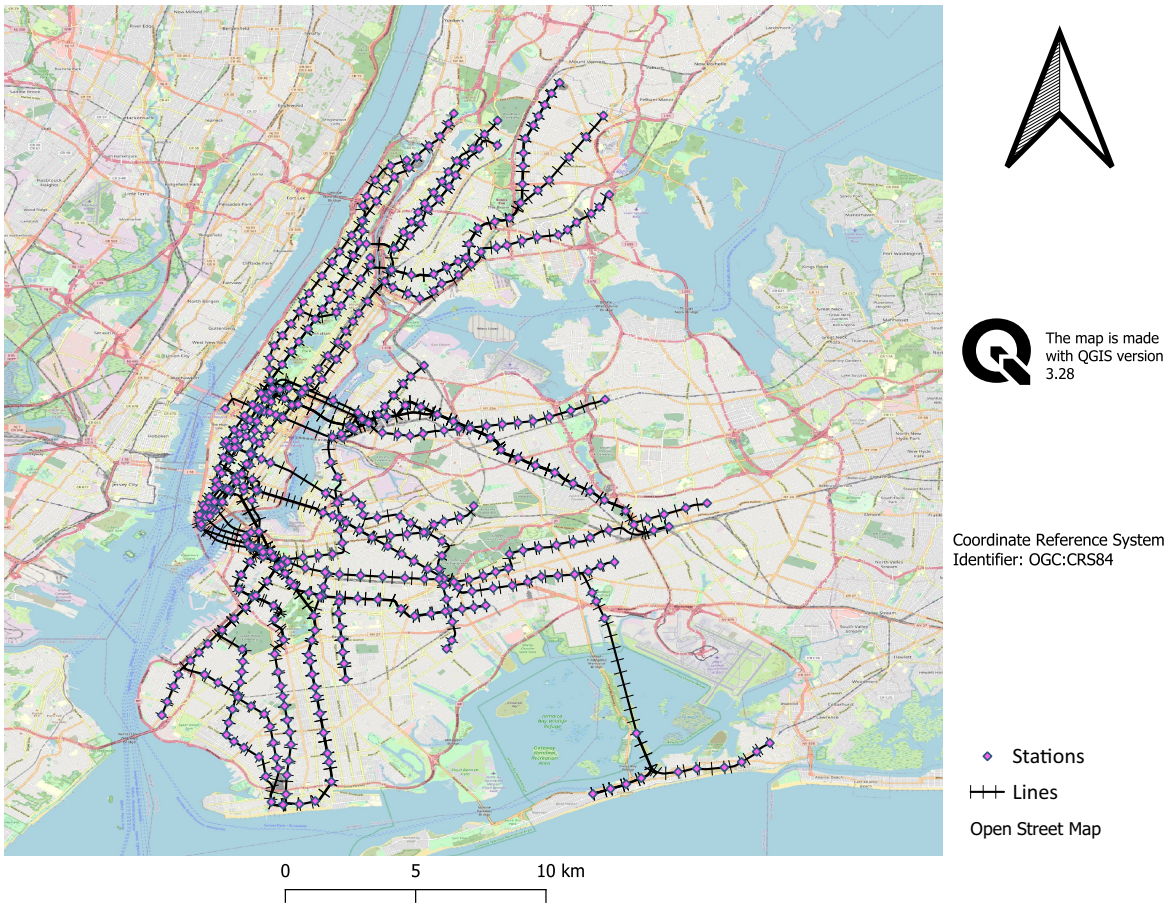
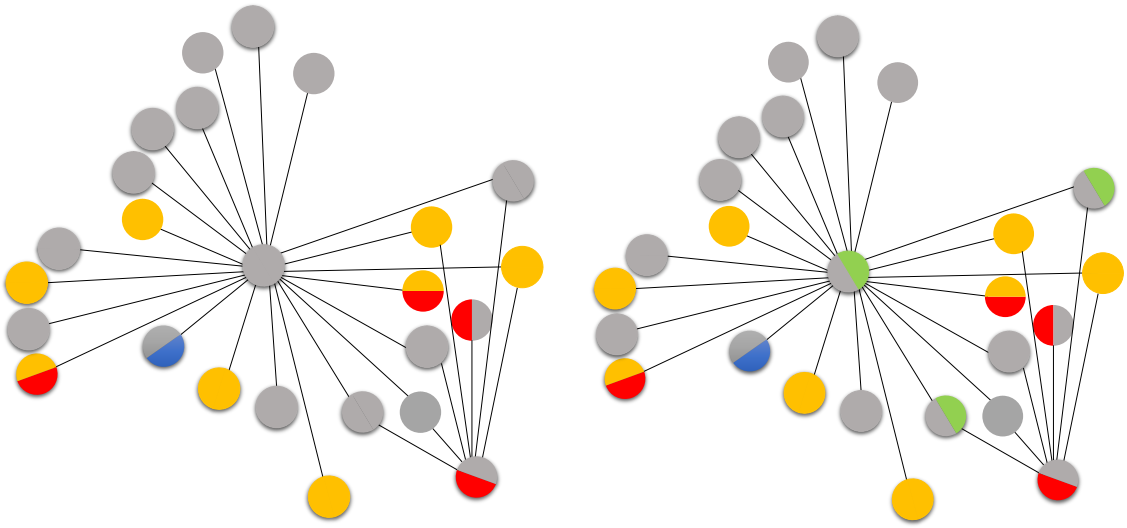


Figure 3.1: New York City Subway network.

closed stations. To illustrate the impact of disinformation interdiction, we close a certain number of stations (removing one node from the subway network and their incoming/outgoing links), ranging from 1 to 20 stations, under the split-up and link-up scenarios in the morning and afternoon. Then, we solve the optimization model and calculate the projected impact of disinformation interdiction from the information layer to the physical layer by determining the delay improvements, defined as the percentage of reduction in average travel time a subway passenger has to travel to get to their destination after engaging with disinformation. Figure 3.4 represents the percentage of delay improvement during disinformation spread in 80 scenarios (i.e., split-up and link-up scenarios in the morning and afternoon by closing a certain number



(a) A set of users before interdiction

(b) A set of users after interdiction

Figure 3.2: Social interaction graph before and after interdiction under the scenario morning split-up.

of stations ranging from 1 to 20). In this figure, the horizontal axis represents the total number of stations that are assumed closed due to disinformation, and the vertical axis represents the percentage of delay improvement found by solving the proposed optimization problem. For example, in Figure 3.4, if five stations are rumored to be closed, we improve the delay in travel time by 5% on average, under the split-up scenario, by interdicting disinformation that the stations marked by red circles in Figure 3.5(a) are closed. However, the delay improvement is substantially larger when more stations are rumored to be closed. If we interdict disinformation regarding the stations rumored to be closed marked by red circles in Figure 3.5(b), we improve delay time by 10% on average. Moreover, interdicting disinformation considerably improves the delay of some exemplary travels, shown with the outlying points in Figure 3.4(c), during the morning rush hour under the link-up scenario starting with the fourth station. This station followed by the next two stations, fifth and sixth, are marked by red circles in Figure 3.6 as a sample of stations rumored to be closed under the

aforementioned scenario. These stations are the Beverly Road, Grand Army Plaza, and Atlantic Av - Barclay's Center, respectively, and interdicting disinformation on closure of these stations reduce delays for some passengers substantially. Since the link-up scenario encourages passengers to reroute and pass through the most trafficked stations, interdicting disinformation about the closure of these stations considerably reduces the delays (over 200% for some passengers) of the travels and stops passengers pass through the crowded stations.

The results of Figure 3.4 further show that if the weaponized disinformation suggests the second most-traveled stations are closed (i.e., a split-up scenario) in the morning, the delay is reduced almost equally up to the 7th most-traveled closed station. Also, under the afternoon split-up scenario, the delay is improved the most as the 11th and 12th most-traveled stations are rumored to be closed due to disinformation. The delay improvement remains constant in other instances. The results of the link-up scenario suggest that interdictions reduce delays as more stations are rumored to be closed by disinformation, specifically in the morning. In addition, disinformation makes some passengers unable to find any route to complete their trip. As the number of rumored closed stations increases to 20 and the subway network is split into disconnected islands, users who traveled from one island to another could not complete their travel since they avoided stations rumored to be closed. We observed a few instances of this issue, 4.5%, 5%, 1.6%, and 12% in the scenarios, morning split-up, afternoon split-up, morning link-up, and afternoon link-up, respectively. The interdiction model helps users complete their travel with minimum travel time and reduces the chance they assume that some stations are closed, making it impossible for them to complete their travel via the subway system.

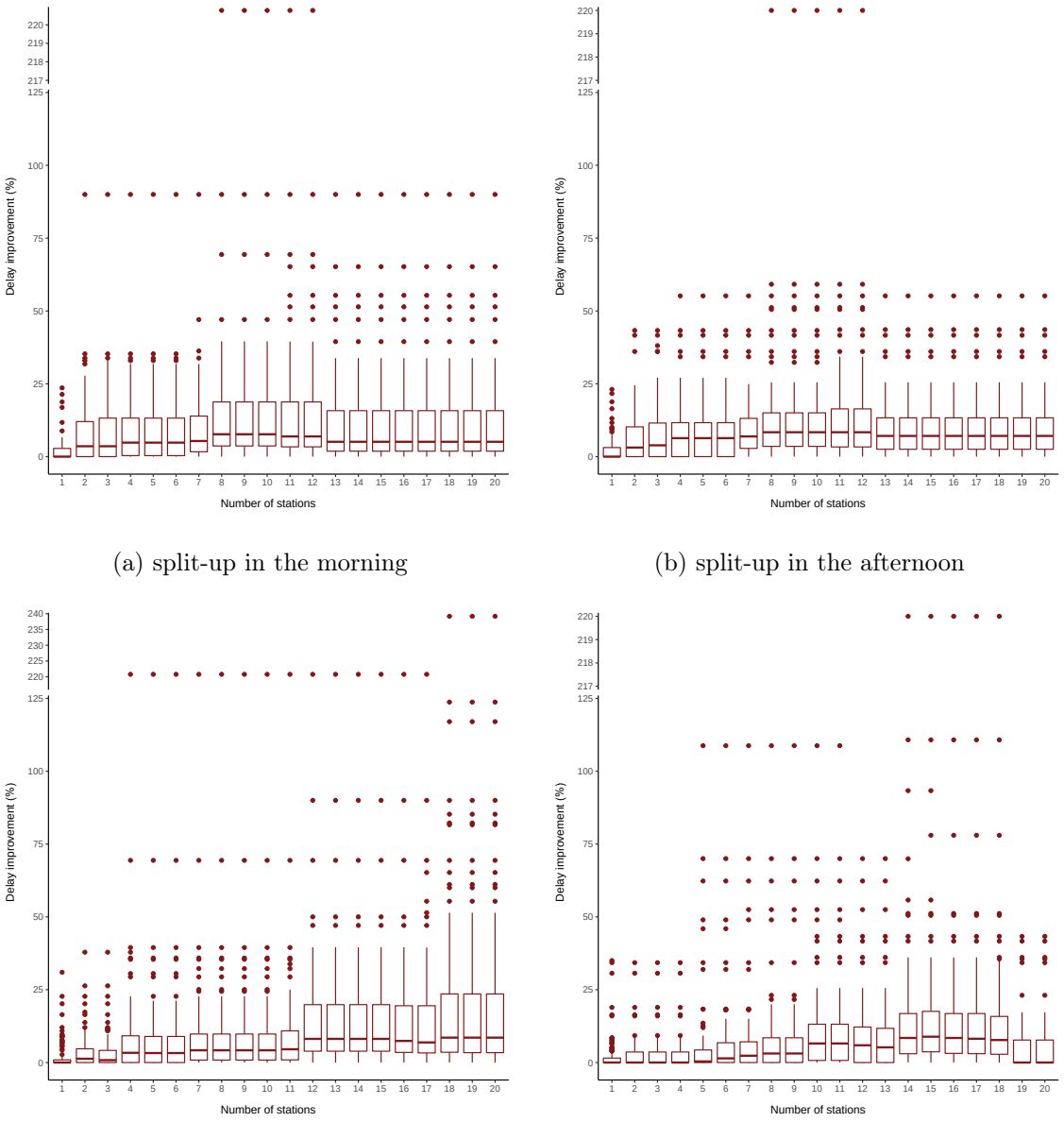


(a) A route not impacted by disinformation



(b) A route impacted by disinformation (reroute)

Figure 3.3: (a) A route not impacted by disinformation; (b) A route impacted by disinformation (reroute), both under the split-up scenario.



(a) split-up in the morning

(b) split-up in the afternoon

(c) link-up in the morning

(d) link-up in the afternoon

Figure 3.4: Percentage of delay improvement (reduction) under scenarios (a) split-up in the morning, (b) split-up in the afternoon, (c) link-up in the morning, and (d) link-up in the afternoon.



(a) The first 5 stations closed by rumor under the split-up scenario



(b) The first 12 stations closed by rumor under the split-up scenario

Figure 3.5: (a) The first 5 stations closed by rumor under the split-up scenario; (b) The first 12 stations closed by rumor under the split-up scenario.



Figure 3.6: Three stations most impacted by interdicting disinformation under the link-up scenario in the morning.

3.5 Concluding Remarks

Taking into account the interdependent relationship between information and physical layers, we proposed and tested an interdiction model aimed at the prevention of weaponized disinformation attacks. To this end, we linearized the nonlinear competitive information cascade model (Carnes et al., 2007) to be able to estimate it efficiently using commercial solvers. To illustrate the outcomes of the proposed model, we solved a large-scale interdiction problem applied to a social network and investigated the effect of this interdiction on public subway system usage in New York City. Results suggest that under the split-up scenario, wherein passengers are moved away from particular stations due to disinformation, the model improves the delays significantly with few stations rumored to be closed compared to the link-up scenario, wherein disinformation is used to steer passengers toward particular stations. As one might expect, the larger the number of rumored station closures, the more important it is to limit the spread of disinformation and interdict the sources of such disinformation. Also, the model is able to improve the delays relatively more during morning travels compared to afternoon travels on average.

Such a model provides insight into how effective communication can be managed during disinformation attacks to protect different infrastructures from being disrupted. A foundational limitation of our model is the cost of computation. We applied our model to a relatively small sample of the social network for computation; however, social interactions are more complex in modeling and optimization computations. Heuristic methods for the proposed model would aid in approximating the optimal solution at a reasonable computation cost, and we will explore such heuristics in the future. Also, the subway data used in this chapter are estimated. Future works can enrich the subway topology data, more effectively informing the structure of the subway network.

In addition, the proposed model can be applied to social networks of different types, such as those with bidirectional and unidirectional links.

Chapter 4

Dealing with Uncertainty in Weaponized Disinformation Attacks and Infrastructure Disruption

4.1 Introduction

Broadcasting false and deceptive information on social networks is an inexpensive and effective way to trigger a crisis, harm public safety and well-being, and undermine public trust in institutions. Disinformation defined as "information falsely characterizing the state of the system, including rumors, factual errors, and attempts at deception" (Floridi, 2005). Disinformation campaigns can take many forms, such as rumors, factual errors, and attempts at deception, and can be spread through various means, such as social media, phone messages, or email. The prevalence of disinformation campaigns is increasing on online platforms (Vosoughi et al., 2018; Allcott et al., 2019), with widespread consequences. For example, a recent study found that a significant proportion of global online citizens, around 86%, can initially be misled by false news stories (Simpson, 2019), highlighting the need to address the problem of disinformation in our society.

Several recent examples of disinformation spread have had substantial adverse consequences. Social media posts suggested that the consumption of a toxic alcohol, methanol, would eradicate the COVID-19 virus, resulting in fatal and non-fatal cases of methanol poisoning (Hassanian-Moghaddam et al., 2020; Apuke and Omar, 2021).

Numerous tweets promoting conspiracies about the manipulation of the 2020 US presidential election sewed discord in public trust in government processes and led to threats to public officials and other forms of violence (Sharma et al., 2022).

Adversaries who seek to attack critical systems, such as infrastructure networks and supply chains, have two potential ways to do so. The first is a direct approach that involves the use of tools such as viruses and ransomware to launch an attack (Sodhi et al., 2012; Cartwright and Cartwright, 2023; Crosignani et al., 2021; Datta and Acton, 2022; Yeboah-Ofori et al., 2021; Kumar and Mallipeddi, 2022). However, adversaries can also attack these systems indirectly by manipulating the consumption behavior of unsuspecting users through weaponized disinformation. Disinformation campaigns can manipulate people to consume goods and services at a higher rate, creating a shortage of these items. Furthermore, false information disseminated through disinformation campaigns can lead to operational failures in physical infrastructure systems, such as electric power failures due to false price messages (Tang et al., 2019a; Raman et al., 2020; Jamalzadeh et al., 2022), and traffic failures due to false road closure messages (Waniek et al., 2021).

Our understanding of the threat posed to infrastructure systems by weaponized disinformation can be visualized in two layers, as depicted in Figure 1.1. The first layer is the information layer, which is made up of social media and other social interactions. The second layer is the physical layer, which represents the different components of a networked infrastructure system, such as the power network or the supply chain. The connection between these layers is the user behavior shaped by the information layer and exhibited by the physical layer. In the case of a weaponized disinformation attack, it originates at the information layer. It triggers human responses, such as changes in consumption, that ultimately alter the performance of the physical layer.

Strategies to prevent such adverse impacts could focus on combating disinformation in the information layer, including (i) empowering social network users to verify and validate online content (Rehm, 2018; IRX; COR), (ii) imposing government regulations to prevent the spread of rumors by social media users (Santos et al., 2022; Zhu and Yang, 2023; Caled and Silva, 2022), (iii) classifying online content by tech companies to inform users of the validity and original sources of content (Hasan et al., 2021; Mheidly and Fares, 2020), and (iv) exposing accurate information to consumers misled by disinformation to make them aware of disinformation and prevent them reacting to it (Jamalzadeh et al., 2022; Waniek et al., 2021). Alternatively, strategies could focus on the physical layer, where physical components and their capacities can be reinforced to tolerate and absorb unanticipated disruptions (e.g., extensive demand by consumers) (Jamalzadeh et al., 2022; Bechmann, 2020; Bliss et al., 2020).

Complicating the optimal allocation of resources to any of the above strategies are the various sources of uncertainty that exist in the two layers of Figure 1.1. Some sources of uncertainty in the information layer include, but are not limited to: (i) the rate at which social network users engage with disinformation, (ii) the rate at which they detect disinformation and act on it, (iii) the impact level of accurate information broadcasting exposed to the users (i.e., accurate information which encourages the users to not engage with disinformation), and (iv) the number of users sensitive to different kinds of disinformation (e.g., rumored price discount). Also, some sources of uncertainty in the physical infrastructure layer incorporate, but are not limited to: (i) the fraction of commodity usage sensitive to price changes, (ii) the extent to which the commodity usage is sensitive to price changes, and (iii) the average usage of commodity per consumer, which varies from one person to another. The uncertain factors, which govern the spread of disinformation in the information layer and impact the performance of the physical layer, raise several different questions before fighting against

disinformation: (i) How the warfare of accurate information against disinformation can be controlled efficiently in information networks to overcome the adversary consequence of disinformation spread on the physical layer? (ii) What is the outcome of making a decision without perfect knowledge of the factors governing this warfare? (iii) How can we efficiently impact the target users (i.e., those who were deceived by disinformation) in the information layer? In summary, does the outcome of our combating decisions match our expectations? Under the uncertain spread of disinformation in the information layer and its projected impact on the physical layers, the answers to these questions are not clear. So what are the best strategies to protect physical layers from disinformation propagated in the information layer with some sort of toleration of uncertainties? To answer these questions, we should take into account the uncertain factors that govern the spread of disinformation in our accurate information sharing decisions and efficiently reach the commodity consumers of the target infrastructure in information networks.

The contribution of this paper is to address the uncertainty in modeling the information and physical layers that can be targeted by an adversary sharing weaponized disinformation. We do so by proposing a robust equivalent of the integrated Epidemiological + Optimization (EPO) model (Jamalzadeh et al., 2022), in which its solutions supercede the solutions generated by the deterministic EPO model under uncertain realization of spread of disinformation in social networks.

The remainder of the paper is as follows. Section 4.2 discusses a summary of relevant research work. Section 4.3 presents the formulations of the proposed robust optimization model. Section 4.4 illustrates the application of the proposed optimization model and compares it with the deterministic model with the Los Angeles County electric power network. Finally, concluding remarks are presented in Section 4.5.

4.2 Background and Literature Review

Infrastructure networks, such as electric power, telecommunications, transportation, water distribution, and gas networks, play a vital role in ensuring individual well-being and economic productivity. Disruption of these systems, whether due to common-cause failure, natural hazards, accidents, or malevolent attacks, can have widespread consequences for other infrastructure networks and communities that depend on them (Barker et al., 2017; Karakoc et al., 2019). One such malevolent attack that has been previously mentioned involves the spread of disinformation. Demand response management, in particular, is a mechanism that is vulnerable to adversarial manipulation, especially for electric power utilities.

Power utilities are increasingly adopting demand response programs to encourage consumers to reduce or change their electricity use during peak periods (Tang et al., 2019a; Lund et al., 2015). As future energy generation scenarios become more prevalent, demand response programs are expected to become even more critical (Raman et al., 2020; Schuitema et al., 2017). Demand response messages are typically sent through text messages, emails, or social media posts and can influence human consumption behavior, potentially spreading disinformation (Jain et al., 2015; Raman et al., 2019; Jamalzadeh et al., 2022). In addition, overloading physical components of the electric power network due to altered consumption behavior could result in electricity shortages in different parts of a utility's coverage area.

The threat of disinformation to infrastructure networks emphasizes protecting their performance against disruption. To this end, Peng et al. (Peng et al., 2020) developed a nonlinear programming model to protect the power network control centers, which monitor and control the power grid system, against cyberattacks by providing redundant servers and disinforming attackers. The proposed model aims to minimize the

probability of system failure and unfilled demand. Ghorbani-Renani et al. (Ghorbani-Renani et al., 2020) developed a trilevel optimization model to balance protection prior to disruption and restoration after disruption of the components of the infrastructure network to optimize the detection of attacks. The results suggest that investments in reinforcing infrastructure network components before disruptions and appropriately assigning recovery resources after disruptions can reduce the vulnerability of the system. Disinformation can indirectly impact infrastructure network operations by manipulating target consumers to consume goods and/or services in a way that damages infrastructure networks. As such, it is crucial to prevent the spread of disinformation from information networks such as social media, radio, and television by raising the consumer's awareness of the malicious consequences of reacting to disinformation.

There are various approaches to modeling the propagation of influence, information, or disinformation, which can help predict the behavior of users. One such approach is the maximization of social influence, where user behavior is influenced by those on their social network, such as following connections, watching television, listening to radio and face-to-face conversations (Kumar et al., 2022; Xu et al., 2014; Lotf et al., 2022; Li et al., 2018). Influence maximization is commonly used in marketing to target specific users to adopt a product. Another set of methods includes information diffusion models, such as *cascade models* and linear *threshold models*. In cascade models, disinformation spreads iteratively with the probability that a user becomes influenced by a neighbor, while in threshold models, users are more likely to take a predetermined action when its surrounding users who take a similar action exceed a threshold (Borodin et al., 2010; Kermani et al., 2017; Clark and Poovendran, 2011; He et al., 2012). A third set of methods is *competitive influence models*, where multiple types of information compete for influence in social networks. Lastly, *compartmental* or *epidemiological models* (Kermack and McKendrick, 1927) are used to predict the spread of information,

where the SIR model is a commonly used model in the context of disinformation spread, dividing users into susceptible, infected and recovered groups to estimate the portion of users who fall into different groups over time.

Information diffusion models have been extended to include multiple competitors, including the multi-cascade spread model by Zarezade et al. (Zarezade et al., 2017) that predicts user adoption to one of the competing behaviors promoted by multiple correlated cascades (i.e., multiple influences by neighborhood users) on social network users. SIR models have also been extended to address such cases of more than one type of spreading mechanism. This includes a competing multivirus SIR model (Zhang et al., 2022; Bichara et al., 2014; Li et al., 2022), which has analogies to the competitive spread of information versus disinformation. Additionally, some information diffusion models also considered personality traits to model how information spreads between users. *Big five* personality traits, including openness, conscientiousness, extraversion, agreeability, and neuroticism, have been shown to influence the way influence is adopted. For example, an individual with a higher level of openness (to experience) are more likely to share their knowledge with others, while those who exhibit neuroticism have a negative influence on sharing their knowledge with others (Lotfi et al., 2016). The rate for each traits might vary from one person to another, due to their social life background, and is concluded by analyzing the surveys. Naturally, such personality traits may govern the adoption of disinformation (Wolverton and Stevens, 2020; Buchanan and Benson, 2019; Buchanan, 2021; Sampat and Raj, 2022; Giachanou et al., 2020; Barman and Conlan, 2021; Burbach et al., 2019).

The accuracy and precision of information diffusion models to estimate adoption during disinformation campaigns depend on the ability of the modeler to evaluate model parameters, such as the rate at which people engage with disinformation and the rate of detection of disinformation. Several variants of SIR models have been developed to

improve the accuracy and precision of SIR models under uncertain evolution of disinformation spread factors. Cho et al. (Cho et al., 2019) developed an opinion model named *Subjective Logic (SL)* to estimate the dynamic of (dis)information adoption during disinformation campaigns. They incorporate factors such as background knowledge and evidence exposed to the social network users. The proposed model enables social network operators to help convince disinformed users and eliminate disinformation. Chai et al. (Chai et al., 2019) developed a stochastic SIR model to incorporate the variant population of social network users. The model supersedes the classic SIR model to represent the real-life (dis)information adoption dynamics in social networks more accurately. Sun et al. (Sun et al., 2021) developed a variant version of the SIR models represented as uncertain differential equations. The model involves a noise into the transmission parameters of the model (e.g., the rate at which rumor spreaders stop spreading it) and can represent the real-world information diffusion processes with a relatively higher accuracy.

In the information diffusion domain, stochastic variants of SIR models are developed to incorporate uncertain factors that drive the spread of disinformation in social networks. As a result, the proposed models can represent the impact of real-life dynamic of social network behavior and (dis)information adoption with relatively higher accuracy and precision during disinformation campaigns. By estimating the evolution of (dis)information adoption under uncertainty, protection decisions become relatively more efficient, such as counter-disinformation activities during disinformation campaigns and lockdown decisions during a pandemic. In SIR models applied to disinformation spread, measuring the rate at which people engage with and detect disinformation can suffer from uncertainty. This uncertainty can impact our ability to make effective decisions about disinformation spread and infrastructure protection decisions. To account for these uncertainties in disinformation campaigns, we propose a robust

optimization model, derived from the initial integrated EPO model (Epidemiological + Optimization) (Jamalzadeh et al., 2022).

4.3 Proposed Robust Model for Network Protection under Uncertain Disinformation Propagation

In this section, we present the deterministic mixed-integer programming model and explain how we convert it to a robust mixed-integer model.

4.3.1 Deterministic Mixed Integer Programming Model

With the notation in Table 4.1, the mixed integer deterministic optimization model developed by Jamalzadeh et al. (Jamalzadeh et al., 2022) is shown in Eqs. 4.1-4.8.

Table 4.1: Model notation.

Notation	Description
Sets	
V	Set of infrastructure network nodes
E	Set of infrastructure network links
T	Set of periods
Parameters	
$\bar{t}_t$	Duration of each period starting from time t to the beginning of its next period
q_{it}	The amount of supply in node $i \in V$ at time $t \in T$
m_{ijt}	Capacity of link from node i to node j at time $t \in T$
p_{it}	Community size surrounding the node i at time $t \in T$
d^c_{it}	Commodity consumption per capita by the community surrounding the node i at time $t \in T$
r^p_{it}	Proportion of commodity consumption of the community surrounding the node $i \in V$ at time $t \in T$ responsive to price shift
ρ_{it}	Estimated sensitivity of commodity consumption of the community surrounding the node $i \in V$ at time $t \in T$ based on the price shift
$\dot{I}_{it}$	Local derivative (change per unit of time interval $\bar{t}_t$) of the proportion of community surrounding node $i \in V$ targeted by disinformation at time $t \in T$
r_{it}	The proportion at which $\dot{I}_{it}$ can be changed by spreading accurate (counter) information for the community surrounding the node $i \in V$ at time $t \in T$
n^p_t	Total number of target locations informed by accurate information at time $t \in T$
Decision variables	
x_{ijt}	The amount of transmitted commodity (flow) from node $i \in V$ to node $j \in V$ at time $t \in T$
h_{it}	Shortage (undersupplied) amount of commodity at node $i \in V$ at time $t \in T$
e_{it}	Excess (oversupplied) amount of commodity at node $i \in V$ at time $t \in T$
d_{it}	Nominal demand of commodity in node $i \in V$ at time $t \in T$
I_{it}	Proportion of community surrounding node $i \in V$ at time $t \in T$ adopted disinformation
g_{it}	$=1$ if accurate information is released for the surrounding community in node $i \in V$ at time $t \in T$; $=0$ otherwise
ν_{it}	The likelihood of which the accurate information being effective as much as r_{it} in node $i \in V$ at time $t \in T$

Eq. (4.1) is the objective function used to penalize the shortage of commodity demand at demand nodes resulting from consumers reacting to disinformation, summed over all nodes of the infrastructure network within a time horizon. Constraint (4.2) balances the inflow, outflow, produced, and consumed commodities in all nodes within a time horizon. This flow balance equation does not become infeasible, since it contains the subtraction of two positive variables, the shortage and excess $h_{it} - e_{it} \in \mathbb{R}$. Constraint (4.3) limits the capacity of the links. Capacity determines the maximum amount of commodity that can be transferred between the nodes through the links. Constraint (4.4) represents the demand response of the consumers who are targeted for disinformation as a multiplier of the baseline demand. The multiplier is determined in terms of the sensitivity of commodity demand to one unit increase in the price of the commodity (that is, *elasticity*), the number of consumers, the portion of users who are deceived by disinformation, and the portion of commodity usage responding to a price change. Constraint (4.5) accounts for the spread of a precise information strategy to reduce the propagation of disinformation. This constraint controls the portion of users deceived by disinformation, at a rate over time, by reaching users with accurate information instead of disinformation. Constraint (4.6) guarantees that if a node is not selected for protection resources (i.e., the release of accurate information surrounding the node), then there is no way in which the node can be spared from a disinformation attack. Constraint (4.7) sets limits on the number of nodes, with a limited budget, for which its surrounding users can be protected by accurate information. Finally, the set of constraints (4.8) defines the nature of the decision variables.

$$\min_{x,h,e,d,I,g,\nu} \sum_{i\in V,\, t\in T} h_{it} \tag{4.1}$$

s.t.

$$\sum_{j\in V:(j,i)\in E} x_{jit} - \sum_{k\in V:(i,k)\in E} x_{ikt} + h_{it} - e_{it} + q_{it} - d_{it} = 0, \qquad \forall i \in V, \forall t \in T, \tag{4.2}$$

(MIP)

$$x_{ijt} \leq m_{ijt}, \qquad \forall i \in V, \forall j \in V, \forall t \in T, \tag{4.3}$$

$$d_{it} = p_{it} d^c_{it}\big\{(1 - I_{it}) + \{I_{it}[r^p_{it}(1 + \rho_{it}) + (1 - r^p_{it})]\}\big\}, \qquad \forall i \in V, \forall t \in T, \tag{4.4}$$

$$I_{it+\bar{t}_t} - I_{it} + \dot{I}_{it+\bar{t}_t}(r_{it}\ \nu_{it} - 1) \leq 0, \qquad \forall i \in V, \forall t \in T \setminus \{|\ T\ |\}, \tag{4.5}$$

$$\nu_{it} \leq g_{it}, \qquad \forall i \in V, \forall t \in T \setminus \{|\ T\ |\}, \tag{4.6}$$

$$\sum_{i\in V} g_{it} \leq n^p_t, \qquad \forall t \in T \setminus \{|\ T\ |\}, \tag{4.7}$$

$$x_{ijt}, h_{it}, e_{it}, d_{it}, I_{it}, \nu_{it} \in \mathbb{R}_{\geq 0}, \qquad g_{it} \in \{0, 1\}. \tag{4.8}$$

4.3.2 Robust Mixed Integer Programming Model with Polytopic Uncertainty Set

Optimization problems with uncertain parameters are related to the uncertainty set of the parameters, which is determined based on the plausible ranges of these parameters (Bertsimas et al., 2011). Therefore, determining the uncertainty set of parameters is the first step in reformulating mathematical optimization problems into robust optimization problems, where we set upper and lower bounds on uncertain parameters based on their plausible values derived from either the historical data of the parameters or

predictive models. Despite stochastic optimization models, robust optimization techniques do not require a distribution of plausible scenarios (Heyman and Sobel, 2004). Instead, the range of plausible scenarios, the uncertainty set, is considered in robust optimization models. There are several different uncertainty modeling techniques in the literature, such as, commonly, ellipsoidal and polyhedral sets (Bertsimas et al., 2016; Jalilvand-Nejad et al., 2016; Gabrel et al., 2014).

We extend the previous mixed-integer deterministic programming model to account for uncertain factors that govern the spread of disinformation in the SIR model: (i) β, which governs the rate at which disinformation spreads, and (ii) γ, which governs the rate at which deceived users recover from and refrain from reacting to disinformation. The uncertain range of these parameters results in the uncertain range for $\dot{I}_{it}$, which represents the derivative of the portion of users deceived by disinformation. Based on the values that β and γ take along with their variation, the parameter $\dot{I}_{it}$ varies within a range (that is, from the lower to the upper bound). Therefore, we found the polytopic uncertainty model to be more appropriate and generated a polytopic uncertainty set with a lower bound $\mathcal{L}_{it}$ and an upper bound $\mathcal{U}_{it}$, of the parameter $\dot{I}_{it}$, and call it $\mathcal{N}_{\dot{I}_{it}}$, such that

$$\mathcal{N}_{\dot{I}_{it}} = \begin{cases} \{\dot{I}_{it} | \mathcal{L}_{it} \leq \dot{I}_{it} \leq \mathcal{U}_{it}, \overline{\mathcal{L}}_{it+\bar{t}_t} \leq \sum_{t'=t}^{t+\bar{t}_t} \dot{I}_{it'} \leq \overline{\mathcal{U}}_{it+\bar{t}_t}\},\ t = t_0, \\ \{\dot{I}_{it} | \mathcal{L}_{it} \leq \dot{I}_{it} \leq \mathcal{U}_{it},\ \overline{\mathcal{L}}_{it} \leq \sum_{t'=t-\bar{t}_t}^{t} \dot{I}_{it'} \leq \overline{\mathcal{U}}_{it}, \overline{\mathcal{L}}_{it+\bar{t}_t} \leq \sum_{t'=t}^{t+\bar{t}_t} \dot{I}_{it'} \leq \overline{\mathcal{U}}_{it+\bar{t}_t}\},\ \forall t \in T \setminus \{t_0, \mid T \mid\}, \end{cases} \tag{4.9}$$

where t_0 is the first time period. In the first time period, the rate of change of the proportion of community targeted by disinformation can vary within the uncertainty bounds. For the rest of time periods, this rate varies based on the added population (e.g., determined by the rate) to the community targeted by disinformation.

Using this uncertainty set, we convert the model (MIP) as follows.

$$\min_{x,h,e,d,I,g,\nu} \sum_{i\in V,\, t\in T} h_{it} \tag{4.10}$$

s.t.

$$\sum_{j\in V:(j,i)\in E} x_{jit} - \sum_{k\in V:(i,k)\in E} x_{ikt} + h_{it} - e_{it} + q_{it} - d_{it} = 0,$$

$$\forall i \in V, \forall t \in T, \tag{4.11}$$

$(R^0.MIP)$

$$x_{ijt} \le m_{ijt}, \qquad \forall i \in V, \forall j \in V, \forall t \in T, \tag{4.12}$$

$$d_{it} = p_{it} d^c_{it}\big\{(1 - I_{it}) + \{I_{it}[r^p_{it}(1 + \rho_{it}) + (1 - r^p_{it})]\}\big\},$$

$$\forall i \in V, \forall t \in T, \tag{4.13}$$

$$I_{it+\bar{t}_t} - I_{it} + \dot{I}_{it+\bar{t}_t}(r_{it}\ \nu_{it} - 1) \le 0,$$

$$\forall i \in V, \forall t \in T \setminus \{|\ T\ |\}, \forall \dot{I}_{it+\bar{t}_t} \in \mathcal{N}_{\dot{I}_{it+\bar{t}_t}}, \tag{4.14}$$

$$\nu_{it} \le g_{it}, \qquad \forall i \in V, \forall t \in T \setminus \{|\ T\ |\}, \tag{4.15}$$

$$\sum_{i\in V} g_{it} \le n^p_t, \qquad \forall t \in T \setminus \{|\ T\ |\}, \tag{4.16}$$

$$x_{ijt}, h_{it}, e_{it}, d_{it}, I_{it}, \nu_{it} \in \mathbb{R}_{\ge 0}, \qquad g_{it} \in \{0, 1\}. \tag{4.17}$$

The objective function 4.10, constraints 4.11-4.13, and constraints 4.15-4.17 remain the same as the model (MIP). However, we convert the constraint 4.14 so that it incorporates the uncertainty set on the parameter $\dot{I}_{it}$. Since $(r_{it}\ \nu_{it} - 1)$ is a non-positive value, we can rewrite the model $(R^0.MIP)$ as follows.

$$\min_{x,h,e,d,I,g,\nu} \sum_{i\in V,\, t\in T} h_{it} \tag{4.18}$$

s.t.

$$\sum_{j\in V:(j,i)\in E} x_{jit} - \sum_{k\in V:(i,k)\in E} x_{ikt} + h_{it} - e_{it} + q_{it} - d_{it} = 0,$$

$$\forall i \in V, \forall t \in T, \tag{4.19}$$

$(R^1.MIP)$

$$x_{ijt} \le m_{ijt}, \qquad \forall i \in V, \forall j \in V, \forall t \in T, \tag{4.20}$$

$$d_{it} = p_{it} d^c_{it}\big\{(1 - I_{it}) + \{I_{it}[r^p_{it}(1 + \rho_{it}) + (1 - r^p_{it})]\}\big\},$$

$$\forall i \in V, \forall t \in T, \tag{4.21}$$

$$\min_{\dot{I}_{it+\bar{t}_t}} \dot{I}_{it+\bar{t}_t}(r_{it}\ \nu_{it} - 1) \le I_{it} - I_{it+\bar{t}_t},$$

$$\forall i \in V, \forall t \in T \setminus \{|\ T\ |\}, \forall \dot{I}_{it+\bar{t}_t} \in \mathcal{N}_{\dot{I}_{i,t+\bar{t}_t}}, \tag{4.22}$$

$$\nu_{it} \le g_{it}, \qquad \forall i \in V, \forall t \in T, \tag{4.23}$$

$$\sum_{i\in V} g_{it} \le n^p_t, \qquad \forall t \in T, \tag{4.24}$$

$$x_{ijt}, h_{it}, e_{it}, d_{it}, I_{it}, \nu_{it} \in \mathbb{R}_{\ge 0}, \qquad g_{it} \in \{0, 1\}. \tag{4.25}$$

The constraint 4.22 is an inner optimization problem in model $(R^1.MIP)$ as follows.

$$\min_{\dot{I}_{it+\bar{t}_t}} \dot{I}_{it+\bar{t}_t}(r_{it}\ \nu_{it} - 1) \tag{4.26}$$

$$\text{s.t.} \tag{4.27}$$

$$\dot{I}_{it+\bar{t}_t} \in \mathcal{N}_{\dot{I}_{it+\bar{t}_t}}. \tag{4.28}$$

By strong duality, the optimal value of the inner optimization problem is equal to the optimal value of its counterpart dual problem (Boyd et al., 2004). The counterpart dual problem for $t = t_0$ is as follows.

$$\max_{\mu_{it},\lambda_{it},\omega_{it+\bar{t}_t},\phi_{it+\bar{t}_t}} (\mu_{it}\mathcal{U}_i - \lambda_{it}\mathcal{L}_{it} + \omega_{it+\bar{t}_t}\overline{\mathcal{U}}_{it+\bar{t}_t} - \phi_{it+\bar{t}_t}\overline{\mathcal{L}}_{it+\bar{t}_t}) \tag{4.29}$$

$$\text{s.t.} \tag{4.30}$$

$$\mu_{it} - \lambda_{it} + \omega_{it+\bar{t}_t} - \phi_{it+\bar{t}_t} = (r_{it}\ \nu_{it} - 1), \tag{4.31}$$

$$\mu_{it}, \lambda_{it}, \omega_{it+\bar{t}_t}, \phi_{it+\bar{t}_t} \in \mathbb{R}_{\geq 0}, \tag{4.32}$$

and the counterpart dual problem $\forall t \in T \setminus \{t_0, \mid T \mid\}$, is as follows.

$$\max_{\mu_{it},\lambda_{it},\omega_{it},\phi_{it},\xi_{it+\bar{t}_t},\kappa_{it+\bar{t}_t}} (\mu_{it}\mathcal{U}_{it} - \lambda_{it}\mathcal{L}_{it} + \omega_{it}\overline{\mathcal{U}}_{it} - \phi_{it}\overline{\mathcal{L}}_{it} + \xi_{it+\bar{t}_t}\overline{\mathcal{U}}_{it+\bar{t}_t} - \kappa_{it+\bar{t}_t}\overline{\mathcal{L}}_{it+\bar{t}_t}) \tag{4.33}$$

$$\text{s.t.} \tag{4.34}$$

$$\mu_{it} - \lambda_{it} + \omega_{it} - \phi_{it} + \xi_{it+\bar{t}_t} - \kappa_{it+\bar{t}_t} = (r_{it}\ \nu_{it} - 1), \tag{4.35}$$

$$\mu_{it}, \lambda_{it}, \omega_{it}, \phi_{it}, \xi_{it+\bar{t}_t}, \kappa_{it+\bar{t}_t} \in \mathbb{R}_{\geq 0}. \tag{4.36}$$

Replacing the derived counterpart dual of constraint 4.22 in model ($R^1.MIP$), we have the following model.

$$\min_{x,h,e,d,I,g,\nu} \sum_{i\in V,\, t\in T} h_{it} \tag{4.37}$$

s.t.

$$\sum_{j\in V:(j,i)\in E} x_{jit} - \sum_{k\in V:(i,k)\in E} x_{ikt} + h_{it} - e_{it} + q_{it} - d_{it} = 0, \qquad \forall i \in V, \forall t \in T, \tag{4.38}$$

$$x_{ijt} \leq m_{ijt}, \qquad \forall i \in V, \forall j \in V, \forall t \in T, \tag{4.39}$$

$$d_{it} = p_{it} d^c_{it} \big\{ (1 - I_{it}) + \{ I_{it} [r^p_{it}(1 + \rho_{it}) + (1 - r^p_{it})] \} \big\}, \qquad \forall i \in V, \forall t \in T, \tag{4.40}$$

$$\left[\begin{array}{c} \max\limits_{\mu_{it},\lambda_{it},\omega_{it+\bar{t}_t},\phi_{it+\bar{t}_t}} (\mu_{it}\mathcal{U}_{it} - \lambda_{it}\mathcal{L}_{it} + \omega_{it+\bar{t}_t}\overline{\mathcal{U}}_{it+\bar{t}_t} - \phi_{it+\bar{t}_t}\overline{\mathcal{L}}_{it+\bar{t}_t}) \\ \mu_{it} - \lambda_{it} + \omega_{it+\bar{t}_t} - \phi_{it+\bar{t}_t} = (r_{it}\ \nu_{it} - 1) \\ \mu_{it}, \lambda_{it}, \omega_{it+\bar{t}_t}, \phi_{it+\bar{t}_t} \in \mathbb{R}_{\geq 0} \end{array} \right] \leq I_{it} - I_{i,t+\bar{t}_t},$$

$$\forall i \in V, t = t_0, \tag{4.41}$$

$$\left[\begin{array}{c} \max,\limits_{\mu_{it},\lambda_{it},\omega_{it},\phi_{it},\xi_{it+\bar{t}_t},\kappa_{it+\bar{t}_t}} (\mu_{it}\mathcal{U}_{it} - \lambda_{it}\mathcal{L}_{it} + \omega_{it}\overline{\mathcal{U}}_{it} - \phi_{it}\overline{\mathcal{L}}_{it} + \xi_{it+\bar{t}_t}\overline{\mathcal{U}}_{it+\bar{t}_t} - \kappa_{it+\bar{t}_t}\overline{\mathcal{L}}_{it+\bar{t}_t}) \\ \mu_{it} - \lambda_{it} + \omega_{it} - \phi_{it} + \xi_{it+\bar{t}_t} - \kappa_{it+\bar{t}_t} = (r_{it}\ \nu_{it} - 1) \\ \mu_{it}, \lambda_{it}, \omega_{it}, \phi_{it}, \xi_{it+\bar{t}_t}, \kappa_{it+\bar{t}_t} \in \mathbb{R}_{\geq 0} \end{array} \right] \leq$$

$$I_{it} - I_{i,t+\bar{t}_t}, \qquad \forall i \in V, \forall t \in T \setminus \{t_0, |\ T\ |\}, \tag{4.42}$$

$$\nu_{it} \leq g_{it}, \qquad \forall i \in V, \forall t \in T, \tag{4.43}$$

$$\sum_{i\in V} g_{it} \leq n^p_t, \qquad \forall t \in T, \tag{4.44}$$

$$x_{ijt}, h_{it}, e_{it}, d_{it}, I_{it}, \nu_{it} \in \mathbb{R}_{\geq 0}, \qquad g_{it} \in \{0, 1\}. \tag{4.45}$$

The maximum values of the inner objective functions in Eqs. 4.41-4.42 are less than or equal to the term $I_{it} - I_{i,t+\bar{t}_t}$, then each objective function must be less than or equal to this term, regardless of the constraints of the inner problem. Therefore, the above is equivalent to the following optimization problem.

$$\min_{x,h,e,d,I,g,\nu,\lambda,\mu,\omega,\phi,\xi,\kappa} \sum_{i\in V,\, t\in T} h_{it} \tag{4.46}$$

s.t.

$$\sum_{j\in V:(j,i)\in E} x_{jit} - \sum_{k\in V:(i,k)\in E} x_{ikt} + h_{it} - e_{it} + q_{it} - d_{it} = 0, \qquad \forall i \in V, \forall t \in T, \tag{4.47}$$

$$x_{ijt} \le m_{ijt}, \qquad \forall i \in V, \forall j \in V, \forall t \in T, \tag{4.48}$$

$$d_{it} = p_{it} d^c_{it}\big\{(1 - I_{it}) + \{I_{it}[r^p_{it}(1 + \rho_{it}) + (1 - r^p_{it})]\}\big\}, \qquad \forall i \in V, \forall t \in T, \tag{4.49}$$

$$(\mu_{it}\mathcal{U}_{it} - \lambda_{it}\mathcal{L}_{it} + \omega_{it+\bar{t}_t}\overline{\mathcal{U}}_{it+\bar{t}_t} - \phi_{it+\bar{t}_t}\overline{\mathcal{L}}_{it+\bar{t}_t}) \le I_{it} - I_{it+\bar{t}_t}, \qquad \forall i \in V, t = t_0, \tag{4.50}$$

$$\mu_{it} - \lambda_{it} + \omega_{it+\bar{t}_t} - \phi_{it+\bar{t}_t} = (r_{it}\ \nu_{it} - 1), \qquad \forall i \in V, t = t_0, \tag{4.51}$$

$$(\mu_{it}\mathcal{U}_{it} - \lambda_{it}\mathcal{L}_{it} + \omega_{it}\overline{\mathcal{U}}_{it} - \phi_{it}\overline{\mathcal{L}}_{it} + \xi_{it+\bar{t}_t}\overline{\mathcal{U}}_{it+\bar{t}_t} - \kappa_{it+\bar{t}_t}\overline{\mathcal{L}}_{it+\bar{t}_t}) \le I_{it} - I_{it+\bar{t}_t},$$
$$\forall i \in V, \forall t \in T \setminus \{t_0, \mid T \mid\}, \tag{4.52}$$

$$\mu_{it} - \lambda_{it} + \omega_{it} - \phi_{it} + \xi_{it+\bar{t}_t} - \kappa_{it+\bar{t}_t} = (r_{it}\ \nu_{it} - 1), \qquad \forall i \in V, \forall t \in T \setminus \{t_0, \mid T \mid\}, \tag{4.53}$$

$$\nu_{it} \le g_{it}, \qquad \forall i \in V, \forall t \in T, \tag{4.54}$$

$$\sum_{i\in V} g_{it} \le n^p_t, \qquad \forall t \in T, \tag{4.55}$$

$$x_{ijt}, h_{it}, e_{it}, d_{it}, I_{it}, \nu_{it}, \lambda_{it}, \mu_{it}, \omega_{it}, \phi_{it}, \xi_{it+\bar{t}_t}, \kappa_{it+\bar{t}_t} \in \mathbb{R}_{\ge 0}, \qquad g_{it} \in \{0, 1\}. \tag{4.56}$$

By optimizing this model, we find the optimal set of nodes to spread accurate information and combat disinformation to minimize its projected malicious impact resulting in unmet demand in the physical layer.

4.4 Case Study: Robust Protection of the Electric Power Network in Disinformation Campaigns

We illustrate the effect of spreading disinformation in the information layer with an electric power network representing the physical layer. Electric power networks are vital to the well-being of communities and to economic productivity. Electric power flow can be modeled as a network that includes nodes and links, so that nodes such as plants, transformers, and demand points are connected to each other by cables that transmit electricity from supply nodes to demand nodes. The amount of electricity that is transmitted from one node to another is called power flow, which is used to satisfy the electrical power demands of the supply nodes (Manfren, 2012). The highest possible electricity carried by a power transmission line is called flow capacity, which might restrict the amount of electric power flowing from supply to demand nodes in balance. The flow in an electric power network can be effectively represented with a network flow optimization model. In representing electric power networks, for example, the model approximates a steady-state DC power flow algorithm (e.g., (LaRocca et al., 2015)) but does not capture AC behavior or more dynamic transient states of the power system. Therefore, the optimization model is optimistic in relation to the behavior of the real-life power system, but it is sufficient with respect to the aim of the proposed model.

When the demand for electric power changes, say due to a false demand response message that encourages users to consume more electricity, then (i) accurate information should be sent to consumers notifying them of the false discount, and (ii) power flows should be adjusted to deliver enough electricity from supplies to the consumers with an accepted quality.

The proposed model is derived from the deterministic version of the EPO model of Jamalzadeh et al. (Jamalzadeh et al., 2022) that involves constraints to ensure the balance between supply and demand, limit transmission line capacity, and disseminate accurate information as a mechanism to combat weaponized disinformation. The previous deterministic model was solved for different scenarios of disinformation spread. It considers different scenarios by setting the parameters (i) β, the rate at which users engage in disinformation, and (ii) γ, the rate at which users detect disinformation.

Solutions to the deterministic model might not be effective in describing unexpected or uncertain outcomes of disinformation spread, say, under the worst-case scenario where disinformation propagates relatively quickly or with a drastic peak. Furthermore, the solutions to the deterministic model could change drastically when the disinformation spread parameters change, for example, when disinformation reaches a broader range of users during a specific period of time (e.g., peak power use). We estimate the number of users who might adopt disinformation by using the SIR model, which has two parameters, β and γ. We design different scenarios for these parameters and calculate their projected influence on the range of the portion of users who adopt disinformation represented by the variable I in the SIR model. These scenarios involve the average and standard deviation (SD) of each parameter. The scenarios determine the lower and upper bound of the users who adopt disinformation over time, represented by $\mathcal{L}$ and $\mathcal{U}$, respectively.

4.4.1 Data and Model Parameters

Electric power transmission lines are designed to transmit enough electricity from the supply to the demand nodes over the network. Since electrical power usage can vary over time, infrastructure designers leave additional capacities to avoid transmission line failure (e.g., overheating) for unanticipated usage based on several factors, such as the length and magnitude of transmission line voltages (Karimi et al., 2018; Mbuli et al., 2019; Kiessling et al., 2003; Begamudre, 2006). The following formula is used to estimate the capacity of transmission lines, where m_{ij} is the capacity of transmission line $ij \in E$, V_{ij} and V_{ji} represent the magnitudes of voltages at the ends of transmission lines ij and ji (kV), respectively, ζ_{ij} is the phase difference between V_{ij} and V_{ji}, ψ is the total positive sequence reactance per phase (ohm per km) and τ_{ij} is the length of the transmission line ij (km).

$$m_{ij} = \frac{V_{ij}V_{ji}sin(\zeta_{ij})}{(\psi\tau_{ij})} \tag{4.57}$$

US state-level data show that a one-unit decrease in the price of electricity can result in a 0.7 unit increase in electricity use, which is known as the price elasticity of the demand for electricity (Miller and Alberini, 2016; Burke and Abayasekara, 2018). It determines the value of ρ in our model, in Eq. 4.49, representing the responsiveness of the use of electrical power to price changes. Although electric power consumption is inelastic to price, the increase in usage is enough to disrupt the electric power network that resulted in shortage.

The response to the use of electric power during disinformation spread can be evaluated by estimating the portion of users who engage with disinformation over time, measured as I in the SIR model. Users engaged in disinformation are assumed to

behave rationally, suggesting that they would take advantage of a discount on electric power and increase electricity usage as much as possible. The portion of the users who engage with disinformation projects on the electric power demand such that (i) the users who adopt disinformation (i.e. false pricing message) are more likely to increase the usage of more variable appliance, such as air conditioner, compared to the non-variable ones, such as refrigerator. We assume that 92% of electricity usage responds to price changes based on the US Energy Information Administration (ENE) and the rest satisfies basic needs, which do not respond to price changes. It determines the value of ρ in our model. (ii) Some users who use social networks may be exposed to disinformation, but not all. We assume that 82% of each population surrounding the electrical power nodes is registered on social networks and is open to receiving the disinformation message, based on the estimated population of direct social network users (STW). This estimation determines the value of r^p in our model.

We use the largest connected synthetic power network (WMN; Soltan et al., 2019) truncated to the boundaries of Los Angeles County (SHP), serving the users surrounding each node within the network. It involves more than 500 nodes, including power generation, demand and transmission nodes, and 700 transmission lines that connect the nodes to each other, as shown in Figure 4.1 (electric power nodes are colored orange and links are colored blue). The publicly available synthetic electric power network data is calculated using *network imitating method based on learning* (NIMBLE) (Soltan et al., 2019). We queried population data from the American Community Survey (ACS) (CEN) and overlayed with geospatial data from the power network in a multi-to-one setting (for example, the population of each census block group is assigned to the closest electric power node).

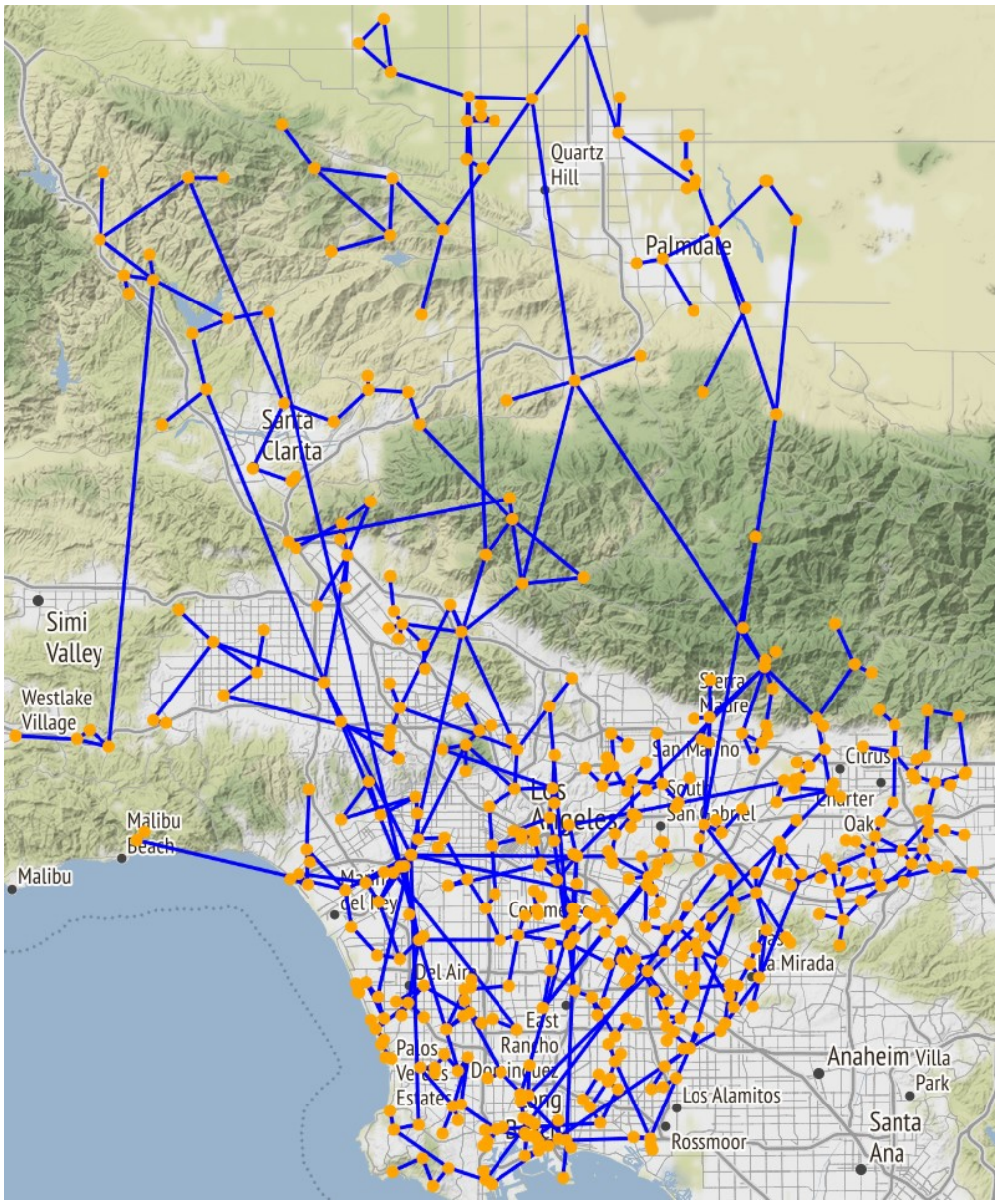


Figure 4.1: Los Angeles County electric power network.

We used the SIR model to estimate the average proportion of users of users who might engage with and react to the disinformation surrounding the electric power nodes. We assume that the disinformation is shared by a social network user and that the rest of the users are potentially susceptible to engage. To solve the model, we generate the upper and lower bounds of the set of polytopic uncertainties based on different average

and standard deviation values of the parameters of the SIR model (i.e., the standard deviation of β and $SD(\gamma)$ are the standard deviation values of β and γ, respectively).

We interpret the inverse parameters of the model, $\frac{1}{\beta}$ and $\frac{1}{\gamma}$, as the time until a contact results in the transmission of disinformation and the time to detect disinformation and correct behavior, respectively. The inherent variability of these parameters results in estimating a range for the rate of change in users who adopt disinformation per unit of time $\dot{I}_{it}$. This range allows us to propose robust infrastructure protection solutions at a relatively higher level of efficiency relative to the deterministic model. We set standard deviations of β and γ, each ranging from 1% to 10% with a step size of 1% and an average of $\frac{1}{\beta}$ and $\frac{1}{\gamma}$ ranging from 1 to 10. We interpret the inverse of parameters, $\frac{1}{\beta}$ and $\frac{1}{\gamma}$, as the average time in which disinformation is transmitted and the average time in which disinformation is detected. A higher standard deviation results in a wider range for $\dot{I}_{it}$. We solved the proposed robust optimization model and compared the solutions with the deterministic model.

4.4.2 Numerical Results

Although the difference between the results of different average values of the parameters, β and γ, is trivial, the solutions vary significantly by different standard deviations of the parameters. Figure 4.2 represents the percentage of improvement in optimal solution by solving the robust optimization model relative to the deterministic optimization model calculated by Eq. 4.58.

$$\frac{\mathcal{Z}_D - \mathcal{Z}_R}{\mathcal{Z}_D} \times 100 \tag{4.58}$$

Here, $\mathcal{Z}_D$ and $\mathcal{Z}_R$ represent the optimal objective value of the deterministic and robust optimization problem, respectively. Figure 4.2 shows that, as the standard deviation of β increases and the standard deviations of γ decrease, the solutions proposed by the robust optimization model are a significant improvement compared to the solutions generated by the deterministic model. This observation indicates that as the range of uncertain users who adopt disinformation is increased, the robust optimization model generates a better solution for the worst-case scenarios.

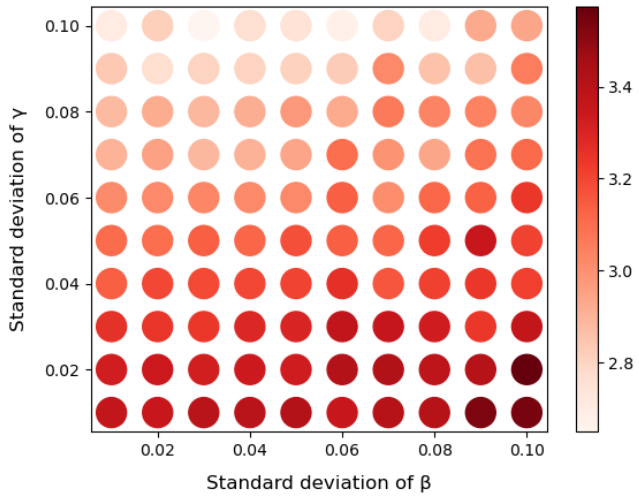
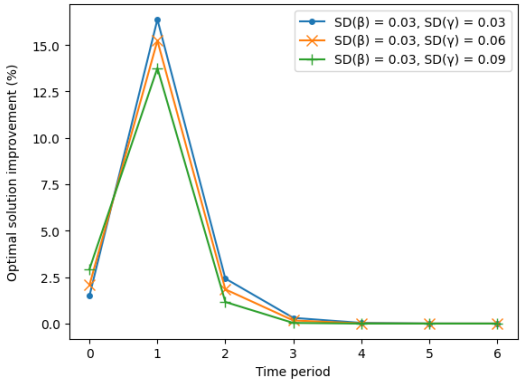


Figure 4.2: The percentage of improvement in the optimal solution by solving the robust optimization model relative to the deterministic optimization problem.

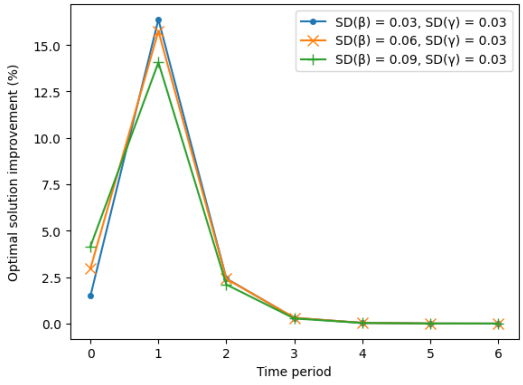
Figure 4.3 represents the percentage of improvement of the optimal solution (i.e., deduction in the electric power shortage) by solving the robust optimization model instead of the deterministic model, over time, for different combinations of standard deviations of β and γ, which represent the rate of user-adopted disinformation and the rate of disinformation detection, respectively. In Figure 4.3 (a), we fix the rate of user-adopted disinformation (that is, the standard deviation of β) and we change the rate of disinformation detection (that is, the standard deviation of γ) to evaluate the sensitivity of the optimal solution with respect to the uncertain outcomes of disinformation spread.

The result reveals that before the number of users adopting disinformation reaches its peak (that is, the highest value during the determined period of time), the higher variability of the rate of disinformation detection results in a better optimal solution in earliest stages, with the robust optimization model. In the later stages, however, solving the robust optimization model with a relatively lower level of uncertainty in the rate of disinformation detection results in a better solutions. Similarly, with a fixed rate of disinformation detection, the optimal solution is improved relatively more under a lower level of uncertainty in the rate of user-adopted disinformation. However, in the later stages, the optimal solution is improved by robust optimization model relatively more with a lower uncertainty level of user-adopted disinformation rate, as shown in Figure 4.3(b). This observation remains similar to the case when we increase the variability of both rates at the same time, as shown in Figure 4.3(c).

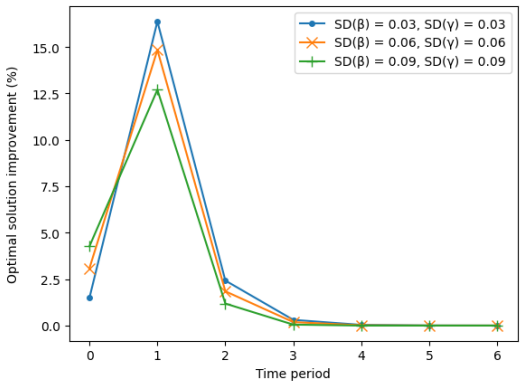
We compare the solutions of target locations informed by accurate information generated by the robust optimization model with the solutions to the deterministic model represented in Figure 4.4. Figure 4.4 (a) shows that, as the variability of the rate of disinformation detection (that is, the standard deviation of γ) decreases, while the variability of the user-adopted disinformation rate is fixed, it is suggested to replace with different locations than the locations advised by the deterministic optimization model. However, changing the rate of user-adopted disinformation, when the rate of disinformation detection is fixed, does not relatively impact the solutions suggested by both robust and deterministic models, as shown in Figure 4.4(b). Similarly to Figure 4.4(a), decreasing both rates causes the robust optimization model to suggest more different solutions (that is, for target locations to share accurate information), with respect to the deterministic model.



(a) Fixed rate of user-adopted disinformation



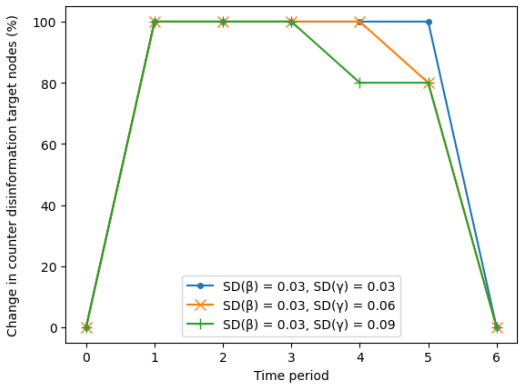
(b) Fixed rate of disinformation detection



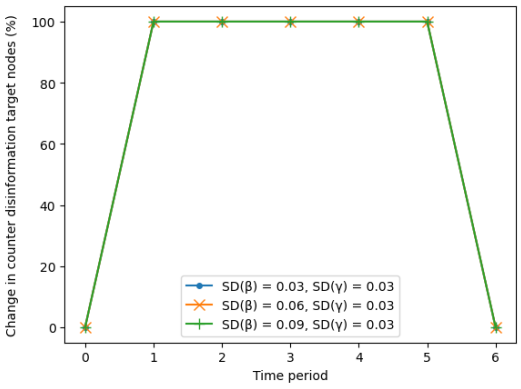
(c) Unfixed rates

Figure 4.3: Optimal improvement of the solution by solving the robust optimization model compared to the deterministic model with (a) fixed rate of user-adopted disinformation and unfixed rate of disinformation detection, (b) fixed rate of disinformation detection and unfixed rate of user-adopted disinformation, and (c) unfixed rates of user-adopted disinformation and disinformation detection, for the combination of parameters $\beta = 1, \gamma = \frac{1}{5}$.

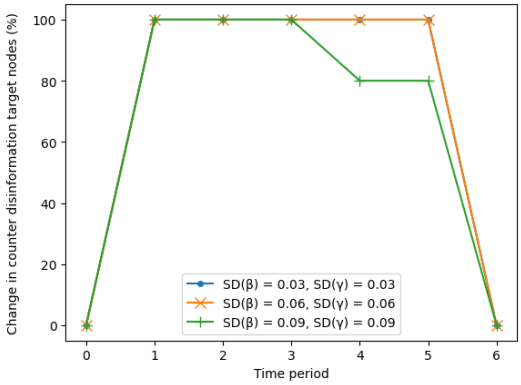
Finally, we compare the optimal solutions suggested by solving the robust optimization model and the deterministic model, over time, for different combinations of the rates of user-adopted disinformation and disinformation detection in Figure 4.5. As shown in Figures 4.5(a)-(c), when we increase the uncertainty levels of both rates at the same time, the optimal solution to the robust optimization problem varies relatively



(a) Fixed rate of user-adopted disinformation



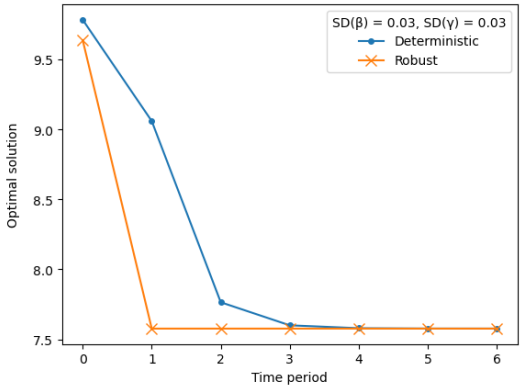
(b) Fixed rate of disinformation detection



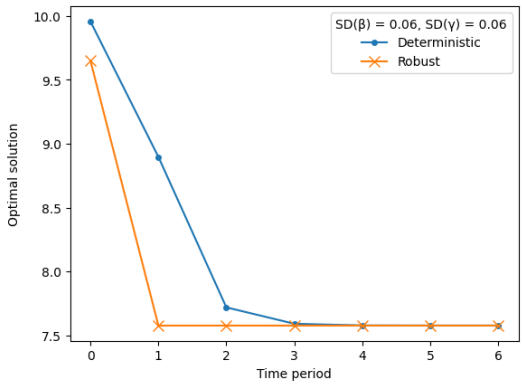
(c) Unfixed rates

Figure 4.4: Change in the average counter information targets by solving the robust optimization model compared to the deterministic model with (a) fixed rate of user-adopted disinformation and unfixed rate of disinformation detection, (b) fixed rate of disinformation detection and unfixed rate of user-adopted disinformation, and (c) unfixed rates of user-adopted disinformation and disinformation detection.

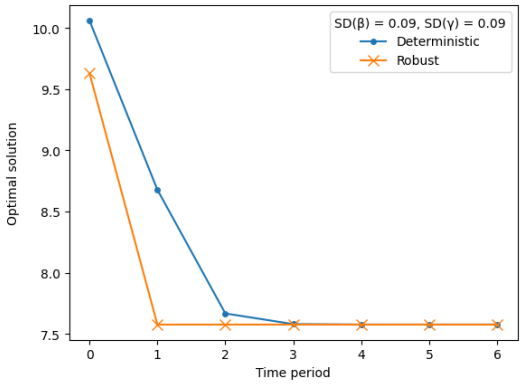
lower, compared to solutions to the deterministic model, that is, the solutions to the robust model is less sensitive to different scenarios of disinformation dissemination.



(a) Fixed rate of user-adopted disinformation



(b) Fixed rate of disinformation detection



(c) Unfixed rates

Figure 4.5: Optimal solutions suggested by solving the robust optimization model compared to the deterministic model with (a) fixed rate of user-adopted disinformation and unfixed rate of disinformation detection, (b) fixed rate of disinformation detection and unfixed rate of user-adopted disinformation, and (c) unfixed rates of user-adopted disinformation and disinformation detection.

4.5 Concluding Remarks

We developed a robust optimization model from the previous version of the infrastructure network protection model against disinformation. Our proposed model in this document aims to propose robust solutions to overcome commodity shortages during disinformation campaigns, which could encourage residents to use commodities in a harmful way with the intention of disrupting commodity distribution systems. The

previous version of the model can propose solutions to mitigate the impact of disinformation campaigns on the performance of the infrastructure network in different scenarios. However, scenarios are not necessarily measurable with high accuracy, so the proposed solutions by solving the deterministic model might not be optimal when the disinformation propagation process changes over time. Our robust optimization model proposed in this paper is capable of handling uncertainty in disinformation propagation, where the proposed solutions can handle uncertainty in estimating the portion of users who might engage with disinformation over a certain period of time. As a result, the solutions to spread accurate information, aimed at combating disinformation, are optimal for a range of uncertainty during disinformation campaigns.

Due to the importance of electric power grids for the economy, we showed the performance of our proposed model for the electric power network in Los Angeles County. However, it can be applied to different infrastructure networks, such as electricity, gas, water, nuclear power plants, and telecommunication networks to combat disinformation campaigns under uncertain propagation of disinformation among residents. Although our proposed robust optimization model involves one commodity (i.e., electric power), it can be developed for (i) multicommodity infrastructure networks and (ii) interdependent infrastructure networks. The former model can consider multiple commodities flowing within a network where the shortage of one commodity impacts the shortage of other commodities due to transport capacities. The latter extension can involve interdependent networks, where the shortage or failure of a commodity in a node over one network can result in the shortage or failure of the interdependent nodes in another network and vice versa. Also, the model involves the combination of both extensions, where each of the interdependent infrastructure networks involves multiple commodities flowing over the network.

We applied the proposed model to the network that includes 500 nodes and 700 links. Although the size of the network was large enough to solve the model in one hour, solving the model for a larger-scale network can become computationally expensive. The solving time might grow exponentially with respect to the topology of the networks being studied. For this reason, we suggest developing methods to solve a large-scale robust network protection model under uncertain spread of disinformation.

There are several different variations of SIR models developed to account for different purposes of (dis)information propagation mechanisms, such as the SIHR model (Zhao et al., 2012b) (that is, H represents Hibernators, defined as the users who might forget or remember (dis)information over time) and the SCIR model (Xiong et al., 2012) (that is, C represents Contacted users who lose their willingness to spread/share the (dis)information as time goes on). In this paper, the SIR model was used to estimate the percentage of social media users who engage in disinformation over time. However, based on different use cases, different variations of SIR models can be used to estimate the range at which social network users adopt disinformation and react to it. Our proposed model is capable of handling different ranges of uncertainty. Further studies might include generating the uncertainty range and solving the model accordingly to protect infrastructure networks from disinformation attacks. Also, any model that can generate the range for uncertain parameters can be deployed as input to our proposed model.

Chapter 5

Conclusions

5.1 Summary of Conclusions

The main goal of this dissertation is to protect infrastructure network from disinformation campaigns. We developed three different strategies to prevent malicious impact of disinformation, propagated in social networks, on physical infrastructure networks.

In the first step, we applied an epidemiological model, the SIR model, to predict the evolution of disinformation propagation over time. The SIR model predicts the fraction of population who engage with disinformation based on different factors such as the rate at which social network users engage with disinformation upon a virtual interaction and the rate at which they detect disinformation and act on it. Eventually, disinformation impacts the behavior, specifically consumption behavior, of the social network users which manifests onto the physical infrastructure network and raises the threat of commodities shortage for the users. We introduced a mechanism to combat against disinformation propagation and its malicious impacts on the physical infrastructure networks. The mechanism involves a subset of population consuming commodities served by different physical infrastructure components (i.e., nodes) associated to the population. Then, we targeted a subset of deceived population (i.e., the population who engaged with disinformation) with accurate information to mitigate the impact of their malicious behavior on the physical infrastructure networks. As a result, we proposed an integrated epidemiological-optimization (EPO) model to predict the fraction

of population who might engage with disinformation over time and expose the deceived population to accurate information as well as balance the physical infrastructure network flows accordingly. We formulated the problem as a mixed integer programming model which integrates a network flow optimization model and an information epidemiological model (i.e., the SIR model). We illustrated the model with the electric power network in Los Angeles County in California and clarified that the proposed model is applicable to other physical infrastructure networks such as water and gas networks. Nonetheless, the applicability of the proposed model to transportation networks requires a different setting compared to electric power, gas, or water networks. This motivated us to propose a mechanism to interdict disinformation and mitigate its malicious impacts on transportation networks in the next step.

In the second step, we proposed a mechanism to interdict the propagation of disinformation by influencing a subset of users and their neighbor connections and mitigate unexpected consumption responses of social network users due to the exposed disinformation. Our proposed mechanism comprises three main stages: (i) we assigned each social network users to a subset of infrastructure network components, say rail transit stations a user pass through, as a one-to-many setting, (ii) we interdict the spread of disinformation by influencing a subset of users and their neighbor connections (i.e., with accurate information), and (iii) we projected the impact of disinformation interdiction mechanism onto the physical infrastructure network. We reformulated a nonlinear integer programming model into a linear mixed integer programming model to be able to estimate solutions to large-scale disinformation interdiction problems within a reasonable resources (i.e., time and computation costs) by commercial solvers using a branch and cut algorithm.

After introducing and developing disinformation counteraction models in the first two steps, the uncertain nature of parameters used in the models motivated us to develop a mathematical optimization model in the third step. The parameters which govern the spread of disinformation might change during the disinformation campaigns, however, the EPO model is solved prior to the disinformation campaigns. So, the physical infrastructure protection solutions might not remain optimal during the disinformation campaign events in real world. Therefore, in the third step, the uncertain nature of the disinformation propagation parameters, the SIR model parameters, motivated us to develop a robust optimization model to tolerate a range of disinformation spread scenarios. As a result, we introduced a robust mixed integer programming model derived from the original EPO model, and showed the solutions to the robust model supercede the solutions to the original model under uncertain realizations of disinformation propagation scenarios.

5.2 Future Directions

This dissertation comprises three main steps of contributions, each of which with limitations and motivations for future research directions.

For the first and third steps, we suggest to solve the EPO model for large-scale problems depending on how the complexity of the model grows with respect to the size of the information and physical infrastructure networks. There are two ways to handle large-scale problems: (i) if the number of variables of the model are large, we suggest to deploy column generation methods, in which, we can solve the problem with limited number of variables then iteratively add new variables of which improve the optimal solution (**?**), and (ii) if the number of constraints are quite large, we suggest to use decomposition methods, where we split the set of constraints into smaller subsets

and solve each of smaller problems either simultaneuously or sequentially (Boyd et al., 2007). Even though we applied the EPO model on the electric power network, the proposed model can also be applied to other infrastructure networks, including water and gas. The variants of SIR models are developed in literature to address different types of social network users assumed to be observed in the population. We recommend to user the variants of the SIR models, such as SIHR (Zhao et al., 2012b) (i.e., H represents the population who forget or remember (dis)information over time) and SCIR (Xiong et al., 2012) models (i.e., C represents the population who lose their willingness to spread (dis)information over time) integrated to the physical infrastructure network flow optimization model. Also, relatively novel models, such as agent-based simulation models, can be used to predict the realization of disinformation campaigns.

To extend our contribution presented as the second step, we similarly suggest to solve the disinformation interdiction problem for large-scale social networks. We illustrated the proposed disinformation interdiction model with a case study involving a subset of Twitter users in information layer, and the New York City subway network. We recommend to apply this model on different social network platforms such as Meta (Facebook) and YouTube, and evaluate the projected impact of disinformation interdiction onto larger-scale infrastructure networks.

Reference List

Barrett, b. 2020. An Artist Used 99 Phones to Fake a Google Maps Traffic Jam. Accessed: 2020-02-01, https://www.wired.com/story/99-phones-fake-google-maps-traffic-jam/.

Can artificial intelligence help end fake news? . Accessed: 2022-05-30, https://ec.europa.eu/research-and-innovation/en/horizon-magazine/can-artificial-intelligence-help-end-fake-news.

DifferentialEquations.jl: Scientific Machine Learning (SciML) Enabled Simulation and Estimation. Accessed: 2022-02-15, https://diffeq.sciml.ai/stable/.

Disinformation about the COVID-19 vaccine is a problem. Stanford researchers are trying to solve it. Accessed: 2022-05-10, https://news.stanford.edu/2022/02/24/curbing-spread-covid-19-vaccine-related-mis-disinformation/.

EIA Website. Accessed: 2021-11-11, https://www.eia.gov/energyexplained/use-of-energy/homes.php.

EIA Website. Accessed: 2023-03-15, https://www.eia.gov/energyexplained/use-of-energy/homes.php.

Fake news is real — A.I. is going to make it much worse. Accessed: 2022-05-30, https://www.cnbc.com/2019/07/12/fake-news-is-real-ai-is-going-to-make-it-much-worse.html.

Fake news spreads faster than true news on Twitter—thanks to people, not bots. Accessed: 2022-05-30, https://www.science.org/content/article/fake-news-spreads-faster-true-news-twitter-thanks-people-not-bots.

Global Online Content Consumption Doubled In 2020. Accessed: 2022-05-30, https://www.forbes.com/sites/johnkoetsier/2020/09/26/global-online-content-consumption-doubled-in-2020/?sh=4d6f6ec12fde.

Gurobi Optimization: The Leader in Decision Intelligence Technology. Accessed: 2023-01-15, https://www.gurobi.com/.

Here's how we're using AI to help detect misinformation. Accessed: 2022-05-30, https://ai.facebook.com/blog/heres-how-were-using-ai-to-help-detect-misinformation/.

How Social Media Moves New York: Twitter Use by Transportation Providers in the New York Region. Accessed: 2022-05-23, https://wagner.nyu.edu/files/faculty/publications/how_social_media_moves_new_york.pdf.

How to combat fake news and disinformation. Accessed: 2022-05-30, https://www.brookings.edu/research/how-to-combat-fake-news-and-disinformation/.

IBM ILOG CPLEX Optimizer. Accessed: 2022-02-15, https://www.ibm.com/analytics/cplex-optimizer.

LA County Boundary Feature Layer. Accessed: 2023-03-15, https://geohub.lacity.org/datasets/lacounty::la-county-boundary-feature-layer/about.

Merriam-Webster Dictionary. Accessed: 2022-01-03, https://www.merriam-webster.com/dictionary/awareness.

Metropolitan Transportation Authority. Accessed: 2022-05-23, https://new.mta.info/coronavirus/ridership.

Misinformation, disinformation, and propaganda: Workshops. Accessed: 2023-02-17, https://guides.library.cornell.edu/evaluate_news.

Number of daily active Snapchat users from 1st quarter 2014 to 1st quarter 2022. Accessed: 2022-05-30, https://www.statista.com/statistics/545967/snapchat-app-dau/.

Number of monetizable daily active Twitter users (mDAU) worldwide from 1st quarter 2017 to 1st quarter 2022 . Accessed: 2022-05-30, https://www.statista.com/statistics/970920/monetizable-daily-active-twitter-users-worldwide/.

Number of monthly active Instagram users from January 2013 to December 2021. Accessed: 2022-05-30, https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/.

NYC Blackout: Subway Passengers Plunged Into Darkness During Manhattan Power Outage. Accessed: 2022-07-04, https://www.cbsnews.com/newyork/news/new-york-city-blackout-subway-loses-power/.

One Year Later: The Texas Freeze Revealed a Fragile Energy System and Inspired Lasting Misinformation. Accessed: 2022-06-06, https://insideclimatenews.org/news/05022022/texas-storms-extreme-weather-renewable-energy/.

QGIS. Accessed: 2022-06-01, https://www.qgis.org/en/site/.

Ransomware Impacting Pipeline Operations. Accessed: 2023-03-27, https://www.cisa.gov/news-events/cybersecurity-advisories/aa20-049a.

Social circles: Twitter. Accessed: 2022-05-24, https://snap.stanford.edu/data/ego-Twitter.html.

Statista Website. Accessed: 2021-12-05, https://www.statista.com/statistics/273476/percentage-of-us-population-with-a-social-network-profile/.

Statista Website. Accessed: 2023-03-15, https://www.statista.com/statistics/273476/percentage-of-us-population-with-a-social-network-profile/.

Stopping disinformation. Accessed: 2023-02-17, https://www.irex.org/stopping-disinformation.

Study: On Twitter, false news travels faster than true stories. Accessed: 2022-05-30, https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308.

Subway Service Guide. Accessed: 2022-05-23, http://web.mta.info/maps/service_guide.pdf.

Texas blackouts fuel false claims about renewable energy. Accessed: 2022-06-06, https://apnews.com/article/false-claims-texas-blackout-wind-turbine-f9e24976e9723021bec21f9a68afe927.

Think You Can Spot Fake News? Many Can't. Accessed: 2023-01-21, https://www.usnews.com/news/health-news/articles/2021-06-01/think-you-can-spot-fake-news-many-cant.

Thomaselli, r., 2020. Man Tries to Delay Flight by Reporting Fake Bomb Threat. Accessed: 2020-02-01, https://www.travelpulse.com/news/airlines/man-tries-to-delay-flight-by-reporting-fake-bomb-threat.html.

Threats to the nation's critical infrastructure. Accessed: 2022-11-07, https://www.cisa.gov/insights.

Three Myths About Renewable Energy and the Grid, Debunked. Accessed: 2022-06-06, https://e360.yale.edu/features/three-myths-about-renewable-energy-and-the-grid-debunked.

Tufnell, n., 2014. Students hack Waze, send in army of traffic bots. Accessed: 2020-02-01, https://www.wired.co.uk/article/waze-hacked-fake-traffic-jam.

Twitter is sweeping out fake accounts like never before, putting user growth at risk. Accessed: 2022-05-30, https://www.washingtonpost.com/technology/2018/07/06/twitter-is-sweeping-out-fake-accounts-like-never-before-putting-user-growth-risk/.

United States Census Bureau. Accessed: 2021-12-13, https://www.census.gov/.

US Census Application Programming Interface. Accessed: 2021-12-11, https://www.census.gov/data/developers/data-sets.html.

US Census Application Programming Interface. Accessed: 2023-03-15, https://www.census.gov/data/developers/data-sets.html.

U.S. Census Bureau. Accessed: 2022-05-23, https://data.census.gov/cedsci/table?q=Population\%20Total&g=1600000US3651000&tid=ACSDT5Y2020.B01003.

Wireless and Mobile Networking Lab. Accessed: 2021-12-07, https://wimnet.ee.columbia.edu/portfolio/synthetic-power-grids-data-sets/.

Wireless and Mobile Networking Lab. Accessed: 2023-03-15, https://wimnet.ee.columbia.edu/portfolio/synthetic-power-grids-data-sets/.

Zetter, K. 2016. Inside the Cunning, Unprecedented Hack of Ukraine's Power Grid. Wired. Accessed: 2021-12-10, https://www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/.

Akella, R., H. Tang, and B. M. McMillin, 2010: Analysis of information flow security in cyber–physical systems. *International Journal of Critical Infrastructure Protection*, **3 (3-4)**, 157–173.

Ali, H., M. S. Khan, A. AlGhadhban, M. Alazmi, A. Alzamil, K. Al-Utaibi, and J. Qadir, 2021: All your fake detector are belong to us: evaluating adversarial robustness of fake-news detectors under black-box settings. *IEEE Access*, **9**, 81 678–81 692.

Allcott, H., M. Gentzkow, and C. Yu, 2019: Trends in the diffusion of misinformation on social media. *Research & Politics*, **6 (2)**, 2053168019848 554.

Almoghathawi, Y., K. Barker, and L. A. Albert, 2019: Resilience-driven restoration model for interdependent infrastructure networks. *Reliability Engineering & System Safety*, **185**, 12–23.

Almoghathawi, Y., A. D. González, and K. Barker, 2021: Exploring recovery strategies for optimal interdependent infrastructure network resilience. *Networks and Spatial Economics*, **21 (1)**, 229–260.

Apuke, O. D., and B. Omar, 2021: Fake news and covid-19: modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*, **56**, 101 475.

Ashrafuzzaman, M., Y. Chakhchoukh, A. A. Jillepalli, P. T. Tosic, D. C. de Leon, F. T. Sheldon, and B. K. Johnson, 2018: Detecting stealthy false data injection attacks in power grids using deep learning. *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*, IEEE, 219–225.

Azucar, D., D. Marengo, and M. Settanni, 2018: Predicting the big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and individual differences*, **124**, 150–159.

Baidya, P. M., W. Sun, and A. Perkins, 2019: A survey on social media to enhance the cyber-physical-social resilience of smart grid. *8th Renewable Power Generation Conference (RPG 2019)*, IET, 1–6.

Banda, M. K., M. Herty, and A. Klar, 2006: Gas flow in pipeline networks. *Networks & Heterogeneous Media*, **1 (1)**, 41.

Barker, K., J. H. Lambert, C. W. Zobel, A. H. Tapia, J. E. Ramirez-Marquez, L. Albert, C. D. Nicholson, and C. Caragea, 2017: Defining resilience analytics for interdependent cyber-physical-social networks. *Sustainable and Resilient Infrastructure*, **2 (2)**, 59–67.

Barman, D., and O. Conlan, 2021: Exploring the links between personality traits and susceptibility to disinformation. *Proceedings of the 32nd ACM Conference on Hypertext and Social Media*, 291–294.

Bechmann, A., 2020: Tackling disinformation and infodemics demands media policy changes. *Digital journalism*, **8 (6)**, 855–863.

Begamudre, R. D., 2006: *Extra high voltage AC transmission engineering*. New Age International.

Ben-Tal, A., L. El Ghaoui, and A. Nemirovski, 2009: *Robust optimization*, Vol. 28. Princeton university press.

Bertsekas, D., 1998: *Network optimization: continuous and discrete models*. Athena Scientific.

Bertsekas, D., 2015: *Convex optimization algorithms*. Athena Scientific.

Bertsimas, D., D. B. Brown, and C. Caramanis, 2011: Theory and applications of robust optimization. *SIAM review*, **53 (3)**, 464–501.

Bertsimas, D., I. Dunning, and M. Lubin, 2016: Reformulation versus cutting-planes for robust optimization: A computational study. *Computational Management Science*, **13**, 195–217.

Beskow, D. M., and K. M. Carley, 2019: Agent based simulation of bot disinformation maneuvers in twitter. 750–761.

Bettencourt, L. M., A. Cintrón-Arias, D. I. Kaiser, and C. Castillo-Chávez, 2006: The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Physica A: Statistical Mechanics and its Applications*, **364**, 513–536.

Bharathi, S., D. Kempe, and M. Salek, 2007: Competitive influence maximization in social networks. *International workshop on web and internet economics*, Springer, 306–311.

Bichara, D., A. Iggidr, and G. Sallet, 2014: Global analysis of multi-strains sis, sir and msir epidemic models. *Journal of Applied Mathematics and Computing*, **44**, 273–292.

Bliss, N., E. Bradley, J. Garland, F. Menczer, S. W. Ruston, K. Starbird, and C. Wiggins, 2020: An agenda for disinformation research. *arXiv preprint arXiv:2012.08572*.

Blume, S. O., F. Corman, and G. Sansavini, 2021: Bayesian origin-destination estimation in networked transit systems using nodal in-and outflow counts. *arXiv preprint arXiv:2105.12798*.

Bodaghi, A., S. Goliaei, and M. Salehi, 2019: The number of followings as an influential factor in rumor spreading. *Applied Mathematics and Computation*, **357**, 167–184.

Borenstein, S., 2009: To what electricity price do consumers respond? residential demand elasticity under increasing-block pricing. *Preliminary Draft April*, **30**, 95.

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

Borodin, A., Y. Filmus, and J. Oren, 2010: Threshold models for competitive influence in social networks. *Internet and Network Economics: 6th International Workshop, WINE 2010, Stanford, CA, USA, December 13-17, 2010. Proceedings 6*, Springer, 539–550.

Boyd, S., S. P. Boyd, and L. Vandenberghe, 2004: *Convex optimization*. Cambridge university press.

Boyd, S., L. Xiao, A. Mutapcic, and J. Mattingley, 2007: Notes on decomposition methods. *Notes for EE364B, Stanford University*, **635**, 1–36.

Bozorgi, A., S. Samet, J. Kwisthout, and T. Wareham, 2017: Community-based influence maximization in social networks under a competitive linear threshold model. *Knowledge-Based Systems*, **134**, 149–158.

Brounen, D., N. Kok, and J. M. Quigley, 2012: Residential energy use and conservation: Economics and demographics. *European Economic Review*, **56 (5)**, 931–945.

Buchanan, T., 2020: Why do people spread false information online? the effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation. *Plos one*, **15 (10)**, e0239 666.

Buchanan, T., 2021: Trust, personality, and belief as determinants of the organic reach of political disinformation on social media. *The Social Science Journal*, 1–12.

Buchanan, T., and V. Benson, 2019: Spreading disinformation on facebook: do trust in message source, risk propensity, or personality affect the organic reach of "fake news"? *Social media+ society*, **5 (4)**, 2056305119888 654.

Burbach, L., P. Halbach, M. Ziefle, and A. Calero Valdez, 2019: Who shares fake news in online social networks? *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, 234–242.

Burke, P. J., and A. Abayasekara, 2018: The price elasticity of electricity demand in the united states: A three-dimensional analysis. *The Energy Journal*, **39 (2)**.

Caled, D., and M. J. Silva, 2022: Digital media and misinformation: An outlook on multidisciplinary strategies against manipulation. *Journal of Computational Social Science*, **5 (1)**, 123–159.

Calo, R., C. Coward, E. S. Spiro, K. Starbird, and J. D. West, 2021: How do you solve a problem like misinformation? *Science advances*, **7 (50)**, eabn0481.

Carnes, T., C. Nagarajan, S. M. Wild, and A. Van Zuylen, 2007: Maximizing influence in a competitive social network: a follower's perspective. *Proceedings of the ninth international conference on Electronic commerce*, 351–360.

Cartwright, A., and E. Cartwright, 2023: The economics of ransomware attacks on integrated supply chain networks. *Digital Threats: Research and Practice*.

Chai, Y., Y. Wang, and L. Zhu, 2019: A stochastic information diffusion model in complex social networks. *IEEE Access*, **7**, 175 897–175 906.

Chen, N., X. Zhu, and Y. Chen, 2019: Information spreading on complex networks with general group distribution. *Physica A: Statistical Mechanics and its Applications*, **523**, 671–676.

Chen, W., Y. Wang, and S. Yang, 2009: Efficient influence maximization in social networks. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 199–208.

Chen, X., 2016: The influences of personality and motivation on the sharing of misinformation on social media. *IConference 2016 Proceedings*.

Cheng, J., Q. Liu, Q. Hui, and F. Choobineh, 2019: The joint optimization of critical interdependent infrastructure of an electricity-water-gas system. 61–73.

Cho, J.-H., S. Rager, J. O'Donovan, S. Adali, and B. D. Horne, 2019: Uncertainty-based false information propagation in social networks. *ACM Transactions on Social Computing*, **2 (2)**, 1–34.

Clack, C., Y. Xie, and A. MacDonald, 2015: Linear programming techniques for developing an optimal electrical system including high-voltage direct-current transmission and storage. *International Journal of Electrical Power & Energy Systems*, **68**, 103–114.

Clark, A., and R. Poovendran, 2011: Maximizing influence in competitive environments: A game-theoretic approach. *Decision and Game Theory for Security: Second International Conference, GameSec 2011, College Park, MD, Maryland, USA, November 14-15, 2011. Proceedings 2*, Springer, 151–162.

Cohen, A. S., L. Lutzke, C. D. Otten, and J. Árvai, 2022: I think, therefore i act: The influence of critical reasoning ability on trust and behavior during the covid-19 pandemic. *Risk Analysis*, **42 (5)**, 1073–1085.

Costa, A., D. Georgiadis, T. S. Ng, and M. Sim, 2018: An optimization model for power grid fortification to maximize attack immunity. *International Journal of Electrical Power & Energy Systems*, **99**, 594–602.

Crosignani, M., M. Macchiavelli, A. F. Silva, and Coauthors, 2021: Cyberattacks and supply chain disruptions. Tech. rep., Federal Reserve Bank of New York.

Csikós, A., T. Charalambous, H. Farhadi, B. Kulcsár, and H. Wymeersch, 2017: Network traffic flow optimization under performance constraints. *Transportation Research Part C: Emerging Technologies*, **83**, 120–133.

Darayi, M., K. Barker, and J. R. Santos, 2017: Component importance measures for multi-industry vulnerability of a freight transportation network. *Networks and spatial economics*, **17 (4)**, 1111–1136.

Datta, P. M., and T. Acton, 2022: From disruption to ransomware: Lessons from hackers. *Journal of Information Technology Teaching Cases*, 20438869221110246.

DeBruhl, B., and P. Tague, 2018: Optimizing a misinformation and misbehavior (mib) attack targeting vehicle platoons. 1–5.

Dormand, J. R., and P. J. Prince, 1980: A family of embedded runge-kutta formulae. *Journal of computational and applied mathematics*, **6 (1)**, 19–26.

Dunning, I., J. Huchette, and M. Lubin, 2017: Jump: A modeling language for mathematical optimization. *SIAM Review*, **59 (2)**, 295–320.

Fang, X., S. Misra, G. Xue, and D. Yang, 2011: Smart grid—the new and improved power grid: A survey. *IEEE communications surveys & tutorials*, **14 (4)**, 944–980.

Fawzi, H., P. Tabuada, and S. Diggavi, 2014: Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic control*, **59 (6)**, 1454–1467.

Fleming, W., A. L. Hayes, K. M. Crosman, and A. Bostrom, 2021: Indiscriminate, irrelevant, and sometimes wrong: Causal misconceptions about climate change. *Risk analysis*, **41 (1)**, 157–178.

Floridi, L., 2005: Is semantic information meaningful data? *Philosophy and phenomenological research*, **70 (2)**, 351–370.

Freeman, L. C., 1977: A set of measures of centrality based on betweenness. *Sociometry*, 35–41.

Gabrel, V., C. Murat, and A. Thiele, 2014: Recent advances in robust optimization: An overview. *European journal of operational research*, **235 (3)**, 471–483.

Garcia Tapia, A., M. Suarez, J. E. Ramirez-Marquez, and K. Barker, 2019: Evaluating and visualizing the economic impact of commercial districts due to an electric power network disruption. *Risk Analysis*, **39 (9)**, 2032–2053.

Ghorbani-Renani, N., A. D. González, K. Barker, and N. Morshedlou, 2020: Protection-interdiction-restoration: Tri-level optimization for enhancing interdependent network resilience. *Reliability Engineering & System Safety*, **199**, 106 907.

Giachanou, A., E. A. Ríssola, B. Ghanem, F. Crestani, and P. Rosso, 2020: The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers. *Natural Language Processing and Information Systems: 25th International Conference on Applications of Natural Language to Information Systems, NLDB 2020, Saarbrücken, Germany, June 24–26, 2020, Proceedings 25*, Springer, 181–192.

Goh, J., S. Adepu, M. Tan, and Z. S. Lee, 2017: Anomaly detection in cyber physical systems using recurrent neural networks. *2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE)*, IEEE, 140–145.

González, A. D., L. Dueñas-Osorio, M. Sánchez-Silva, and A. L. Medaglia, 2016: The interdependent network design problem for optimal infrastructure system restoration. *Computer-Aided Civil and Infrastructure Engineering*, **31 (5)**, 334–350.

Granovetter, M., 1978: Threshold models of collective behavior. *American journal of sociology*, **83 (6)**, 1420–1443.

Hamill, J. T., 2006: *Analysis of layered social networks.* Air Force Institute of Technology.

Hamill, J. T., R. F. Deckro, V. D. Wiley, and R. S. Renfro, 2007: Gains, losses and thresholds of influence in social networks. *International Journal of Operational Research*, **2 (4)**, 357–379.

Han, Q., H. Wen, and F. Miao, 2018: Rumor spreading in interdependent social networks. *Peer-to-Peer Networking and Applications*, **11**, 955–965.

Haraguchi, M., and S. Kim, 2016: Critical infrastructure interdependence in new york city during hurricane sandy. *International Journal of Disaster Resilience in the Built Environment*.

Hasan, M. R., S. Deng, N. Sultana, and M. Z. Hossain, 2021: The applicability of blockchain technology in healthcare contexts to contain covid-19 challenges. *Library Hi Tech*, **39 (3)**, 814–833.

Hassanian-Moghaddam, H., N. Zamani, A.-A. Kolahi, R. McDonald, and K. E. Hovda, 2020: Double trouble: methanol outbreak in the wake of the covid-19 pandemic in iran—a cross-sectional assessment. *Critical Care*, **24 (1)**, 1–3.

He, X., G. Song, W. Chen, and Q. Jiang, 2012: Influence blocking maximization in social networks under the competitive linear threshold model. *Proceedings of the 2012 siam international conference on data mining*, SIAM, 463–474.

He, Z., Z. Cai, and X. Wang, 2015: Modeling propagation dynamics and developing optimized countermeasures for rumor spreading in online social networks. 205–214.

Hethcote, H. W., 2000: The mathematics of infectious diseases. *SIAM review*, **42 (4)**, 599–653.

Heyman, D. P., and M. J. Sobel, 2004: *Stochastic models in operations research: stochastic optimization*, Vol. 2. Courier Corporation.

Horne, B. D., J. Nørregaard, and S. Adali, 2019: Robust fake news detection over time and attack. *ACM Transactions on Intelligent Systems and Technology (TIST)*, **11 (1)**, 1–23.

Hosseini, S., K. Barker, and J. E. Ramirez-Marquez, 2016: A review of definitions and measures of system resilience. *Reliability Engineering & System Safety*, **145**, 47–61.

Hsu, N.-S., and K.-W. Cheng, 2002: Network flow optimization model for basin-scale water supply planning. *Journal of water resources planning and management*, **128 (2)**, 102–112.

Huang, K., C. Zhou, Y. Qin, and W. Tu, 2019: A game-theoretic approach to cross-layer security decision-making in industrial cyber-physical systems. *IEEE Transactions on Industrial Electronics*, **67 (3)**, 2371–2379.

Hunt, K., P. Agarwal, and J. Zhuang, 2022: Monitoring misinformation on twitter during crisis events: a machine learning approach. *Risk analysis*, **42 (8)**, 1728–1748.

Indu, V., and S. M. Thampi, 2020: A systematic review on the influence of user personality in rumor and misinformation propagation through social networks. *International Symposium on Signal Processing and Intelligent Recognition Systems*, Springer, 216–242.

Jain, M., V. Chandan, M. Minou, G. Thanos, T. K. Wijaya, A. Lindt, and A. Gylling, 2015: Methodologies for effective demand response messaging. 453–458.

Jalilvand-Nejad, A., R. Shafaei, and H. Shahriari, 2016: Robust optimization under correlated polyhedral uncertainty set. *Computers & Industrial Engineering*, **92**, 82–94.

Jamalzadeh, S., K. Barker, A. D. González, and S. Radhakrishnan, 2022: Protecting infrastructure performance from disinformation attacks. *Scientific Reports*, **12 (1)**, 1–14.

Jin, F., E. Dougherty, P. Saraf, Y. Cao, and N. Ramakrishnan, 2013: Epidemiological modeling of news and rumors on twitter. 1–9.

Karakoc, D. B., Y. Almoghathawi, K. Barker, A. D. González, and S. Mohebbi, 2019: Community resilience-driven restoration model for interdependent infrastructure networks. *International Journal of Disaster Risk Reduction*, **38**, 101 228.

Karakoc, D. B., K. Barker, and A. D. González, 2023: Analyzing the tradeoff between vulnerability and recoverability investments for interdependent infrastructure networks. *Socio-Economic Planning Sciences*, 101508.

Karimi, S., P. Musilek, and A. M. Knight, 2018: Dynamic thermal rating of transmission lines: A review. *Renewable and Sustainable Energy Reviews*, **91**, 600–612.

Kaufman, S. M., C. Qing, N. Levenson, M. Hanson, and Coauthors, 2012: Transportation during and after hurricane sandy.

Keeling, M. J., and K. T. Eames, 2005: Networks and epidemic models. *Journal of the royal society interface*, **2 (4)**, 295–307.

Kempe, D., J. Kleinberg, and É. Tardos, 2003: Maximizing the spread of influence through a social network. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 137–146.

Kermack, W. O., and A. G. McKendrick, 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.

Kermani, M. A. M. A., S. F. F. Ardestani, A. Aliahmadi, and F. Barzinpour, 2017: A novel game theoretic approach for modeling competitive information diffusion in social networks with heterogeneous nodes. *Physica A: statistical mechanics and its applications*, **466**, 570–582.

Khurana, P., and D. Kumar, 2018: Sir model for fake news spreading through whatsapp. 26–27.

Kiessling, F., P. Nefzger, J. F. Nolasco, and U. Kaintzyk, 2003: *Overhead power lines: planning, design, construction*, Vol. 759. Springer.

Kovendan, A., and D. Sridharan, 2017: Development of smart grid system in india: a survey. 275–285.

Kumar, S., A. Mallik, A. Khetarpal, and B. Panda, 2022: Influence maximization in social networks using graph embedding and graph neural network. *Information Sciences*, **607**, 1617–1636.

Kumar, S., and R. R. Mallipeddi, 2022: Impact of cybersecurity on operations and supply chain management: Emerging trends and future research directions. *Production and Operations Management*, **31 (12)**, 4488–4500.

Lai, K., X. Xiong, X. Jiang, M. Sun, and L. He, 2020: Who falls for rumor? influence of personality traits on false rumor belief. *Personality and Individual Differences*, **152**, 109 520.

Lambiotte, R., and M. Kosinski, 2014: Tracking the digital footprints of personality. *Proceedings of the IEEE*, **102 (12)**, 1934–1939.

LaRocca, S., J. Johansson, H. Hassel, and S. Guikema, 2015: Topological performance measures as surrogates for physical flow models for risk and vulnerability analysis for electric power systems. *Risk analysis*, **35 (4)**, 608–623.

Larson, H. J., L. Lin, and R. Goble, 2022: Vaccines and the social amplification of risk. *Risk Analysis*.

Leng, J., Q. Guo, B. Ma, S. Zhang, and P. Sun, 2020: Bridging personality and online prosocial behavior: The roles of empathy, moral identity, and social self-efficacy. *Frontiers in Psychology*, 2436.

Leskovec, J., and J. Mcauley, 2012: Learning to discover social circles in ego networks. *Advances in neural information processing systems*, **25**.

Leuthold, F. U., H. Weigt, and C. von Hirschhausen, 2012: A large-scale spatial optimization model of the european electricity market. *Networks and spatial economics*, **12 (1)**, 75–107.

Li, B., K. Barker, and G. Sansavini, 2017: Measuring community and multi-industry impacts of cascading failures in power systems. *IEEE Systems Journal*, **12 (4)**, 3585–3596.

Li, C., J. Wang, J. Xu, and Y. Rong, 2022: The global dynamics of a sir model considering competitions among multiple strains in patchy environments. *Mathematical Biosciences and Engineering*, **19 (5)**, 4690–4702.

Li, M., H. Wang, and H. Wang, 2019: Resilience assessment and optimization for urban rail transit networks: A case study of beijing subway network. *IEEE Access*, **7**, 71 221–71 234.

Li, W., Y. Ding, Y. Yang, R. S. Sherratt, J. H. Park, and J. Wang, 2020: Parameterized algorithms of fundamental np-hard problems: a survey. *Human-Centric Computing and Information Sciences*, **10 (1)**, 1–24.

Li, Y., J. Fan, Y. Wang, and K.-L. Tan, 2018: Influence maximization on social graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*, **30 (10)**, 1852–1872.

Liang, G., W. He, C. Xu, L. Chen, and J. Zeng, 2015: Rumor identification in microblogging systems based on users' behavior. *IEEE Transactions on Computational Social Systems*, **2 (3)**, 99–108.

Liang, G., J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, 2016: A review of false data injection attacks against modern power systems. *IEEE Transactions on Smart Grid*, **8 (4)**, 1630–1638.

Liao, T.-Y., T.-Y. Hu, and Y.-N. Ko, 2018: A resilience optimization model for transportation networks under disasters. *Natural hazards*, **93**, 469–489.

Liu, Q., T. Li, and M. Sun, 2017: The analysis of an seir rumor propagation model on heterogeneous network. *Physica A: Statistical Mechanics and its Applications*, **469**, 372–380.

Liu, W., and Z. Song, 2020: Review of studies on the resilience of urban critical infrastructure networks. *Reliability Engineering & System Safety*, **193**, 106 617.

Lobban, H., Y. Almoghathawi, N. Morshedlou, and K. Barker, 2021: Community vulnerability perspective on robust protection planning in interdependent infrastructure networks. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, **235 (5)**, 798–813.

Lotf, J. J., M. A. Azgomi, and M. R. E. Dishabi, 2022: An improved influence maximization method for social networks based on genetic algorithm. *Physica A: Statistical Mechanics and its Applications*, **586**, 126 480.

Lotfi, M., S. N. B. Muktar, A. C. Ologbo, and K. C. Chiemeke, 2016: The influence of the big-five personality traits dimensions on knowledge sharing behavior. *Mediterranean Journal of Social Sciences*, **7 (1 S1)**, 241.

Luenberger, D. G., 1997: *Optimization by vector space methods*. John Wiley & Sons.

Lund, P. D., J. Lindgren, J. Mikkola, and J. Salpakari, 2015: Review of energy system flexibility measures to enable high levels of variable renewable electricity. *Renewable and sustainable energy reviews*, **45**, 785–807.

Lyons, B. A., J. M. Montgomery, A. M. Guess, B. Nyhan, and J. Reifler, 2021: Overconfidence in news judgments is associated with false news susceptibility. *Proceedings of the National Academy of Sciences*, **118 (23)**.

Mahabub, A., 2020: A robust technique of fake news detection using ensemble voting classifier and comparison with other classifiers. *SN Applied Sciences*, **2 (4)**, 525.

Manfren, M., 2012: Multi-commodity network flow models for dynamic energy management–mathematical formulation. *Energy Procedia*, **14**, 1380–1385.

Martin, A., M. Möller, and S. Moritz, 2006: Mixed integer models for the stationary case of gas network optimization. *Mathematical programming*, **105 (2)**, 563–582.

Mbuli, N., R. Xezile, L. Motsoeneng, M. Ntuli, and J.-H. Pretorius, 2019: A literature review on capacity uprate of transmission lines: 2008 to 2018. *Electric Power Systems Research*, **170**, 215–221.

McCormick, G. P., 1976: Computability of global solutions to factorable nonconvex programs: Part i—convex underestimating problems. *Mathematical programming*, **10 (1)**, 147–175.

Medvedeva, M., T. E. Simos, C. Tsitouras, and V. Katsikis, 2021: Direct estimation of sir model parameters through second-order finite differences. *Mathematical Methods in the Applied Sciences*, **44 (5)**, 3819–3826.

Meel, P., and D. K. Vishwakarma, 2020: Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications*, **153**, 112 986.

Mheidly, N., and J. Fares, 2020: Leveraging media and health communication strategies to overcome the covid-19 infodemic. *Journal of public health policy*, **41 (4)**, 410–420.

Miller, M., and A. Alberini, 2016: Sensitivity of price elasticity of demand to aggregation, unobserved heterogeneity, price trends, and price endogeneity: Evidence from us data. *Energy Policy*, **97**, 235–249.

Mishra, S., X. Li, T. Pan, A. Kuhnle, M. T. Thai, and J. Seo, 2016: Price modification attack and protection scheme in smart grid. *IEEE Transactions on Smart Grid*, **8 (4)**, 1864–1875.

Molina, M. D., and S. S. Sundar, 2019: Technological affordances can promote misinformation. *Journalism and Truth in an Age of Social Media*, 40–57.

Najarian, M., and G. J. Lim, 2020: Optimizing infrastructure resilience under budgetary constraint. *Reliability Engineering & System Safety*, **198**, 106 801.

Nasrolahpour, E., H. Ghasemi, and M. Khanabadi, 2012: Optimal transmission congestion management by means of substation reconfiguration. 416–421.

Nguyen, L. N., J. D. Smith, and M. T. Thai, 2019: Vulnerability assessment of social-smart grids: An algorithmic approach. 1–7.

Ortiz-Ospina, E., and M. Roser, 2023: The rise of social media. *Our world in data*.

Ouyang, M., 2014: Review on modeling and simulation of interdependent critical infrastructure systems. *Reliability engineering & System safety*, **121**, 43–60.

Pastor-Satorras, R., C. Castellano, P. Van Mieghem, and A. Vespignani, 2015: Epidemic processes in complex networks. *Reviews of modern physics*, **87 (3)**, 925.

Peng, R., H. Xiao, J. Guo, and C. Lin, 2020: Defending a parallel system against a strategic attacker with redundancy, protection and disinformation. *Reliability Engineering & System Safety*, **193**, 106 651.

Qi, J., X. Liang, Y. Wang, and H. Cheng, 2018: Discrete time information diffusion in online social networks: micro and macro perspectives. *Scientific reports*, **8 (1)**, 1–15.

Qiang, Z., E. L. Pasiliao, and Q. P. Zheng, 2019: Model-based learning of information diffusion in social media networks. *Applied Network Science*, **4 (1)**, 1–16.

Raman, G., B. AlShebli, M. Waniek, T. Rahwan, and J. C.-H. Peng, 2020: How weaponizing disinformation can bring down a city's power grid. *PloS one*, **15 (8)**, e0236 517.

Raman, G., J. C.-H. Peng, and T. Rahwan, 2019: Manipulating residents' behavior to attack the urban power distribution system. *IEEE Transactions on Industrial Informatics*, **15 (10)**, 5575–5587.

Rehm, G., 2018: An infrastructure for empowering internet users to handle fake news and other online media phenomena. *Language Technologies for the Challenges of the Digital Age: 27th International Conference, GSCL 2017, Berlin, Germany, September 13-14, 2017, Proceedings 27*, Springer, 216–231.

Rocco, C. M., K. Barker, J. Moronta, and J. E. Ramirez-Marquez, 2018: Multiobjective formulation for protection allocation in interdependent infrastructure networks using an attack-diffusion model. *Journal of Infrastructure Systems*, **24 (1)**, 04018 002.

Rodrigues, H. S., 2016: Application of sir epidemiological model: new trends. *arXiv preprint arXiv:1611.02565*.

Rui, X., F. Meng, Z. Wang, G. Yuan, and C. Du, 2018: Spir: The potential spreaders involved sir model for information diffusion in social networks. *Physica A: Statistical Mechanics and its Applications*, **506**, 254–269.

Sahafizadeh, E., and B. T. Ladani, 2018: The impact of group propagation on rumor spreading in mobile social networks. *Physica A: Statistical Mechanics and its Applications*, **506**, 412–423.

Sampat, B., and S. Raj, 2022: Fake or real news? understanding the gratifications and personality traits of individuals sharing fake news on social media platforms. *Aslib Journal of Information Management*.

Santos, M. L. R., M. C. Paim, C. L. M. Soares, D. M. Santos, R. S. Sande, and G. R. d. M. Santos, 2022: Government actions to address the disinformation crisis during the covid-19 pandemic. *Saúde em Debate*, **45**, 187–204.

Santos-D'amorim, K., and M. K. F. de Oliveira Miranda, 2021: Misinformation, disinformation, and malinformation: Clarifying the definitions and examples in disinfodemic times. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, **26**, 01–23.

Schreiber, D., C. Picus, D. Fischinger, and M. Boyer, 2021: The defalsif-ai project: protecting critical infrastructures against disinformation and fake news. *e & i Elektrotechnik und Informationstechnik*, **138 (7)**, 480–484.

Schuitema, G., L. Ryan, and C. Aravena, 2017: The consumer's role in flexible energy systems: An interdisciplinary approach to changing consumers' behavior. *IEEE Power and Energy Magazine*, **15 (1)**, 53–60.

Sharma, K., E. Ferrara, and Y. Liu, 2022: Characterizing online engagement with disinformation and conspiracies in the 2020 us presidential election. *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16, 908–919.

Shrivastava, G., P. Kumar, R. P. Ojha, P. K. Srivastava, S. Mohan, and G. Srivastava, 2020: Defensive modeling of fake news through online social networks. *IEEE Transactions on Computational Social Systems*, **7**, 1159–1167.

Simpson, S., 2019: Fake news: A global epidemic, vast majority (86%) of online global citizens have been exposed to it. *Ipsos. June*, **11**.

Sodhi, M. S., C. S. Tang, M. S. Sodhi, and C. S. Tang, 2012: Supply chain risk management. *Managing supply chain risk*, 3–11.

Soltan, S., A. Loh, and G. Zussman, 2019: A learning-based method for generating synthetic power grids. *IEEE Systems Journal*, **13 (1)**, 625–634.

Staiano, J., B. Lepri, N. Aharony, F. Pianesi, N. Sebe, and A. Pentland, 2012: Friends don't lie: inferring personality traits from social network structure. *Proceedings of the 2012 ACM conference on ubiquitous computing*, 321–330.

Sun, H., Y. Sheng, and Q. Cui, 2021: An uncertain sir rumor spreading model. *Advances in Difference Equations*, **2021 (1)**, 286.

Super, J. F., P. Li, G. Ishqaidef, and J. P. Guthrie, 2016: Group rewards, group composition and information sharing: A motivated information processing perspective. *Organizational Behavior and Human Decision Processes*, **134**, 31–44.

Tahiri, A., D. Ladeveze, P. Chiron, B. Archimede, and L. Lhuissier, 2018: Reservoir management using a network flow optimization model considering quadratic convex cost functions on arcs. *Water Resources Management*, **32 (10)**, 3505–3518.

Tang, D., Y.-P. Fang, E. Zio, and J. E. Ramirez-Marquez, 2019a: Resilience of smart power grids to false pricing attacks in the social network. *IEEE Access*, **7**, 80 491–80 505.

Tang, D., Y. P. Fang, E. Zio, and J. E. Ramirez-Marquez, 2019b: Resilience of smart power grids to false pricing attacks in the social network. *IEEE Access*, **7**, 80 491–80 505.

Tang, J., H. Zhu, and J. Guo, 2022: Information diffusion between users in open data ecosystem: Modelling and simulation analysis. *Mathematical Problems in Engineering*, **2022**.

Tian, W., X. Ji, W. Liu, G. Liu, J. Zhai, Y. Dai, and S. Huang, 2020: Prospect theoretic study of honeypot defense against advanced persistent threats in power grid. *IEEE Access*, **8**, 64 075–64 085.

Tobler, W. R., 1969: Geographical filters and their inverses. *Geographical Analysis*, **1 (3)**, 234–253.

Treen, K. M. d., H. T. Williams, and S. J. O'Neill, 2020: Online misinformation about climate change. *Wiley Interdisciplinary Reviews: Climate Change*, **11 (5)**, e665.

Tsitouras, C., 2011: Runge–kutta pairs of order 5 (4) satisfying only the first column simplifying assumption. *Computers & Mathematics with Applications*, **62 (2)**, 770–775.

Vasin, A., O. Grigoryeva, and N. Tsyganov, 2020: A model for optimization of transport infrastructure for some homogeneous goods markets. *Journal of Global Optimization*, **76 (3)**, 499–518.

Vosoughi, S., D. Roy, and S. Aral, 2018: The spread of true and false news online. *Science*, **359 (6380)**, 1146–1151.

Wang, B., Z. Yuan, X. Liu, Y. Sun, B. Zhang, and Z. Wang, 2021a: Electricity price and habits: Which would affect household electricity consumption? *Energy and Buildings*, **240**, 110 888.

Wang, Q., Z. Lin, Y. Jin, S. Cheng, and T. Yang, 2015: Esis: emotion-based spreader–ignorant–stifler model for information diffusion. *Knowledge-based systems*, **81**, 46–55.

Wang, Y., F. Qing, J. P. Chai, and Y. P. Ni, 2021b: Spreading dynamics of a 2sih2r, rumor spreading model in the homogeneous network. *Complexity*, **2021**.

Wang, Y., J. Wang, H. Wang, R. Zhang, and M. Li, 2021c: Users' mobility enhances information diffusion in online social networks. *Information Sciences*, **546**, 329–348.

Waniek, M., G. Raman, B. AlShebli, J. C.-H. Peng, and T. Rahwan, 2021: Traffic networks are vulnerable to disinformation attacks. *Scientific reports*, **11 (1)**, 1–11.

Wardle, C., H. Derakhshan, and Coauthors, 2018: Thinking about "information disorder": formats of misinformation, disinformation, and mal-information. *Ireton, Cherilyn; Posetti, Julie. Journalism,"fake news"& disinformation. Paris: Unesco*, 43–54.

Watts, D. J., D. M. Rothschild, and M. Mobius, 2021: Measuring the news and its impact on democracy. *Proceedings of the National Academy of Sciences*, **118 (15)**, e1912443 118.

Wirtz, M., M. Hahn, T. Schreiber, and D. Müller, 2021: Design optimization of multi-energy systems using mixed-integer linear programming: Which model complexity and level of detail is sufficient? *Energy Conversion and Management*, **240**, 114 249.

Wirz, C. D., M. A. Xenos, D. Brossard, D. Scheufele, J. H. Chung, and L. Massarani, 2018: Rethinking social amplification of risk: Social media and zika in three languages. *Risk Analysis*, **38 (12)**, 2599–2624.

Wolverton, C., and D. Stevens, 2019: The impact of personality in recognizing disinformation. *Online Information Review*.

Wolverton, C., and D. Stevens, 2020: The impact of personality in recognizing disinformation. *Online information review*, **44 (1)**, 181–191.

Wong, J. C. S., J. Z. Yang, Z. Liu, D. Lee, and Z. Yue, 2021: Fast and frugal: Information processing related to the coronavirus pandemic. *Risk Analysis*, **41 (5)**, 771–786.

Woo, J., and H. Chen, 2016: Epidemic model for information diffusion in web forums: experiments in marketing exchange and political dialog. *SpringerPlus*, **5 (1)**, 1–19.

Woo, J., J. Son, and H. Chen, 2011: An sir model for violent topic diffusion in social media. 15–19.

Xiong, F., Y. Liu, Z.-j. Zhang, J. Zhu, and Y. Zhang, 2012: An information diffusion model based on retweeting mechanism for online social media. *Physics letters A*, **376 (30-31)**, 2103–2108.

Xu, W., Z. Lu, W. Wu, and Z. Chen, 2014: A novel approach to online social influence maximization. *Social Network Analysis and Mining*, **4**, 1–13.

Yeboah-Ofori, A., S. Islam, S. W. Lee, Z. U. Shamszaman, K. Muhammad, M. Altaf, and M. S. Al-Rakhami, 2021: Cyber threat predictive analytics for improving cyber supply chain security. *IEEE Access*, **9**, 94 318–94 337.

Zarezade, A., A. Khodadadi, M. Farajtabar, H. Rabiee, and H. Zha, 2017: Correlated cascades: Compete or cooperate. *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31.

Zhang, C., S. Gracy, T. Başar, and P. E. Paré, 2022: A networked competitive multi-virus sir model: Analysis and observability. *IFAC-PapersOnLine*, **55 (13)**, 13–18.

Zhang, Z., S. Radhakrishnan, C. Subramanian, K. Barker, and A. D. González, 2021: Causal node failures and computation of giant and small components in networks. *IEEE Transactions on Network Science and Engineering*, **8 (4)**, 3048–3060.

Zhao, L., J. Wang, Y. Chen, Q. Wang, J. Cheng, and H. Cui, 2012a: Sihr rumor spreading model in social networks. *Physica A: Statistical Mechanics and its Applications*, **391**, 2444–2453.

Zhao, L., J. Wang, Y. Chen, Q. Wang, J. Cheng, and H. Cui, 2012b: Sihr rumor spreading model in social networks. *Physica A: Statistical Mechanics and its Applications*, **391 (7)**, 2444–2453.

Zhao, L., J. Wang, and R. Huang, 2015: Immunization against the spread of rumors in homogenous networks. *PloS one*, **10 (5)**, e0124 978.

Zhao, Z., and Coauthors, 2020: Fake news propagates differently from real news even at early stages of spreading. *EPJ Data Science*, **9 (1)**, 1–14.

Zhu, X., and S. Yang, 2023: 1 toward a sociotechnical framework for misinformation policy analysis. *The Usage and Impact of ICTs during the Covid-19 Pandemic*.