2021

Modeling Coupled Disease-Behavior Dynamics of SARS-CoV-2 Using Influence Networks

Juliana C. Taube
Bowdoin College

Modeling Coupled Disease-Behavior Dynamics of SARS-CoV-2 Using Influence Networks

An Honors Paper for the Department of Mathematics

Juliana C. Taube

Bowdoin College, 2021

Abstract

SARS-CoV-2, the virus that causes COVID-19, has caused significant human morbidity and mortality since its emergence in late 2019. Not only have over three million people died, but humans have been forced to change their behavior in a variety of ways, including limiting their contacts, social distancing, and wearing masks. Early infectious disease models, like the classical SIR model by Kermack and McKendrick, do not account for differing contact structures and behavior. More recent work has demonstrated that contact structures and behavior can considerably impact disease dynamics. We construct a coupled disease-behavior dynamical model for SARS-CoV-2 by incorporating heterogeneous contact structures and decisions about masking. We use a contact network with household, work, and friend interactions to capture the variation in contact patterns. We allow decisions about masking to occur at a different time scale from disease spread which dramatically changes the masking dynamics. Drawing from the field of game theory, we construct an individual decision-making process that relies on perceived risk of infection, social influence, and individual resistance to masking. Through simulation, we find that social influence prevents masking, while perceived risk largely drives individuals to mask. Underlying contact structure also affects the number of people who mask. This model serves as a starting point for future work which could explore the relative importance of social influence and perceived risk in human decision-making.

Acknowledgements

To my advisors, Professors Mary Lou Zeeman and Mohammad Irfan, I am deeply appreciative of your guidance and support this past year. With your help, I have grown as a researcher and have learned more than I could have imagined about how to be a good scientist. I feel so lucky to have worked with you both on this project. Thank you.

To the faculty in the Bowdoin Math Department, thank you for teaching me how to be a problem-solver. I have so enjoyed the challenging, yet collaborative nature of all of my math classes here at Bowdoin. Thank you especially to Professor Barker for sharing your passion for mathematics and helping me see that I could be a math major. To Professor Taback, thank you for always lending a listening ear and for your thoughtful advising; taking MATH 2020 with you will always be one of my most memorable experiences at Bowdoin. Thank you to Professor O'Brien for sharing your enthusiasm for the world of probability and statistics. I hope there will be more Bayesian statistics in my future.

Thank you also to Dj Merrill for help navigating the HPC.

To my friends, thank you for listening to me ramble about the bugs in my code and any new updates on my project. I appreciate your positivity and interest more than you know.

And finally, to my parents, who supported me from the very beginning of this journey all the way to the end. I could not have done this without you. Thank you for your endless support, encouragement to do what makes me happy, and for being the most incredible role models. I love you both so much.

Contents

Abstract **i**

Acknowledgements **ii**

List of Figures **v**

List of Tables **xiii**

1 Introduction **1**

1.1 Background 1

1.2 Our Contribution 3

1.2.1 Progression of Disease 4

1.2.2 Heterogeneous Contact Patterns 4

1.2.3 Shifting Behaviors 5

2 Disease Progression **8**

2.1 Compartmental Models 8

2.2 Our Disease Model 11

3 Networks and Contact Structures **15**

3.1 Networks Overview 15

3.2 Our Network 19

3.3 Disease Spread on Networks 22

4 Behavior and Influence **23**

4.1 Disease-Behavior Models 23

4.2 Game Theory 25

4.3 Influence Games 29

4.4 Our Behavior Model 31

5 Modeling Results **35**

5.1 Parameters 35

5.2 Network Characteristics 37

5.3 Model Simulations 39

5.3.1 No Masking Behavior 40

5.3.2 Mask Mandate 40

5.3.3 Role of Resistance 42

5.3.4 Role of Perceived Risk 50

5.3.5 Role of Social Influence 53

5.3.6 Role of Equilibration 59

5.3.7 Role of Network Structure 59

6 Conclusions **65**

6.1 Key Findings 65

6.2 Future Work 67

List of Figures

2.1 Movement of individuals through SIR compartments. 9

2.2 Progression of infection and disease via symptomatic or asymptomatic routes in our model. The E compartment is important for the time delay between infection and infectiousness. Distinction between infectiousness states I_1, I_2, A_1 and A_2 allows for differentiation of infectiousness in these stages, as indicated by preliminary data [49, 27, 7]. 12

2.3 Varying probabilities of transmission given contact between infectious (orange and red) and susceptible (blue) individuals depending on symptoms and state of infectiousness of infectious individual. 13

2.4 Number of time steps (days) that individuals spend in the exposed and infectious compartments. 14

3.1 Example of an undirected five node network. The blue and yellow circles are nodes, and the black lines connecting them are edges. Node 0 has a degree of 4, while nodes 1 and 4 have a degree of 3, and nodes 2 and 3 have a degree of 2. 16

3.2 In highly clustered networks, nodes with mutual friends (e.g., B and C) are also likely to be connected. In networks with less clustering, nodes with mutual friends are less likely to also be friends. 17

3.3 Multilayer network, with three layers: household, workplace, friend group. Each layer contains the same nodes, but the connections differ between layers and are assigned different edge weights depending on the layer. 20

3.4 Distribution of household sizes based on US ACS Census data from 2018. All households with seven or more people are considered to have exactly seven people. 21

4.1 Prisoner's Dilemma two-player game theory example. Suspect 1 and 2 must each decide whether or not to confess to a crime they committed. The payoffs – the prison sentences – are given as Suspect 1, Suspect 2. The Nash equilibrium outcome, where both players simultaneously play their best response and have no incentive to deviate, is shown in blue. 26

5.1 Degree distributions of four networks of one thousand nodes generated using the multilayer method outlined in Chapter 3. (A) network 1, (B) network 2, (C) network 3, (D) network 4. 37

5.2 Degree distribution of Erdos-Renyi random networks of one thousand nodes generated with $p = 0.014$. (A) Erdos-Renyi network 1, (B) Erdos-Renyi network 2. 38

5.3 Degree distribution of Barabasi-Albert preferential attachment network of one thousand nodes generated with $m = 7$. (A) Barbasi-Albert network 1, (B) Barbasi-Albert network 2. 38

5.4 No masking behavior. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. 41

5.5 No masking behavior. Outbreak statistics for each multilayer network. (A) Total number of nodes infected. (B) Peak number of daily new infections. (C) Time step of peak number of daily new infections. 41

5.6 Mask mandate with 85% compliance. Outbreak statistics for each multilayer network. (A) Total number of nodes infected. (B) Peak number of daily new infections. (C) Time step of peak number of daily new infections. 42

5.7 No resistance to masking; only perceived risk and social influence contribute to masking decision. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval. 43

5.8 No resistance to masking; only perceived risk and social influence contribute to masking decision. Outbreak statistics for each multilayer network. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections. 44

5.9 Varying maximum resistance to masking. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each column is a maximum resistance level, 0.01, 0.1, and 0.2, from left to right. Each row is a network; network 1, 2, 3, and 4 from top to bottom. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval. 45

5.10 Varying maximum resistance to masking. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each column is a maximum resistance level, 0.4, 0.8, and 1, from left to right. Each row is a network; network 1, 2, 3, and 4 from top to bottom. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval. 46

5.11 Varying maximum resistance to masking. Outbreak statistics for each multilayer network. Each column is a maximum resistance level, 0.01, 0.1, and 0.2, from left to right. Each row is a statistic. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections. 47

5.12 Varying maximum resistance to masking. Outbreak statistics for each multilayer network. Each column is a maximum resistance level, 0.4, 0.8, and 1, from left to right. Each row is a statistic. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections. 48

5.13 Maximum resistance of 0.2. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval. 49

5.14 Maximum resistance of 0.2. Outbreak statistics for each multilayer network. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections. 49

5.15 No perceived risk of infection; only resistance (up to 0.2) and social influence contribute to masking decision. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes. 50

5.16 No perceived risk of infection; only resistance (up to 0.2) and social influence contribute to masking decision. Outbreak statistics for each multilayer network. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections. 51

5.17 Perceived risk incorporating proportion of masking neighbors. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval. 52

5.18 Perceived risk incorporating proportion of masking neighbors. Outbreak statistics for each multilayer network. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections. 53

5.19 No social influence; only resistance (up to 0.2) and perceived risk contribute to masking decision. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval. 54

5.20 No social influence; only resistance (up to 0.2) and perceived risk contribute to masking decision. Outbreak statistics for each multilayer network. (A) Peak number of daily new infections. (B) Time step of peak number of daily new infections. (C) Total number of nodes infected. (D) Total number of nodes who masked at some point during the outbreak. 55

5.21 No social influence; only resistance (up to 0.2) and perceived risk contribute to masking decision. Number of masking nodes per day shown for 10 different simulations on multilayer network 1 (1000 nodes). Each simulation is a different color. Note that the light purple, blue, and green curves appear to have multiple waves of increased number of maskers per day. 55

5.22 Varying maximum incoming social influence. Each column is a maximum influence level, 0.1, 0.25, 0.5, 0.75, and 1, from left to right. Each row is a network; network 1, 2, 3, and 4 from top to bottom. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval. 57

5.23 Varying maximum incoming social influence. Outbreak statistics for each multilayer network. Each column is a maximum influence level, 0.1, 0.25, 0.5, 0.75, and 1, from left to right. Each row is a statistic. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections. 58

5.24 No equilibration in the decision-making process. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval. 60

5.25 No equilibration in the decision-making process. Outbreak statistics for each multilayer network. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections. 60

5.26 Two rounds of equilibration in the decision-making process. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval. 61

5.27 Two rounds of equilibration in the decision-making process. Outbreak statistics for each multilayer network. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections. 61

5.28 Number of newly infected nodes per time step for two one thousand node Erdos-Renyi random networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval. . 62

5.29 Outbreak statistics for two one thousand node Erdos-Renyi random networks. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections. 63

5.30 Number of newly infected nodes per time step for two one thousand node Barabasi-Albert preferential attachment networks. Columns are seeded differently; randomly, highest degree node, lower degree node, from left to right. Each row is a different network. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval. . 64

5.31 Outbreak statistics for each Barabasi-Albert preferential attachment network. Networks are either seeded randomly, with the highest degree node, or with a lower degree node. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections. 64

6.1 Summary of results. Ten simulations conducted on multilayer network 1 with 1000 nodes with varying behavior models. Note that Barabasi-Albert and Erdos-Renyi simulations were conducted on the respective networks. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections. 66

List of Tables

5.1 Parameter list. 36

Chapter 1

Introduction

1.1 Background

As of May 20, 2021, coronavirus disease-2019 (COVID-19) has taken the lives of over 3.4 million people worldwide, and nearly 588,000 in the US alone [110, 34]. Moreover, nearly 165 million people have been infected with SARS-CoV-2, the virus that causes COVID-19, some suffering long term health consequences [32, 34, 5, 98, 22]. COVID-19 is not the first emerging disease to cause substantial human morbidity and mortality; influenza, HIV/AIDS, and SARS epidemics have also demonstrated the devastation that can result from novel pathogens [78]. Not only is it hard to predict where and when a new pathogen will emerge, but after emergence, it takes time to characterize the pathogen and determine what measures will prevent its spread.

In January of 2020, when newspapers first reported on a mysterious pneumonia in China [107], little was understood about SARS-CoV-2. Policy-makers and scientists were considering whether the novel virus would cause an epidemic [96, 63]. With evidence of human-to-human transmission and increasing cases in Iran, Italy, and South Korea in March 2020, it became clear that COVID-19 would cause more than an epidemic but a pandemic [109]. Researchers and policy-makers began deliberating over how to slow the spread of disease:

Which non-pharmaceutical interventions (NPIs) should be implemented and when? Experts now agree that masks, social distancing, and avoiding gatherings can slow the spread of COVID-19, but early in the pandemic the route of disease transmission was less certain [23, 47]. Moreover, research is still being conducted to understand when it is appropriate to put each measure in place. Trade-offs between loss of life and loss of freedom and/or income must be weighed in decisions about lockdown or stay-at-home orders [6, 101, 103]. Although many aspects of SARS-CoV-2 are still not well understood, substantial progress has been made since January 2020.

Reaching this level of comprehension was not simple [51]. Initial estimates of epidemic size and intervention effectiveness depended on unreliable information about the contagiousness of SARS-CoV-2, number of cases, whom it infects, how it is transmitted, contact patterns of its human hosts, and how human behavior would evolve over time [38]. With so many factors at work coupled to large amounts of uncertainty, many in the science community turned to mathematical models to predict the extent of viral spread and to assess potential interventions. These models could indicate general trends resulting from different possible parameters or preventive measures, ultimately informing key policy decisions.

Mathematical models have long been used to study infectious disease dynamics, starting in the early 1900s. One of the earliest and simplest infectious disease models is the compartmental SIR model introduced by Kermack and McKendrick in 1927 [58]. In this model, individuals move through three consecutive compartments (S, I, and R) over time. These compartments represent the course of disease; initially, individuals are Susceptible, then they become Infectious, and eventually Removed as they recover or die from the disease [58]. The rate at which individuals move from one compartment to the next is mainly determined by the transmissibility of the pathogen (governing movement from S to I) and the recovery rate (governing movement from I to R). Kermack and McKendrick implemented the model using ordinary differential equations (ODEs). As all models must make assumptions about the biological system of interest, the classic SIR model assumes that: immunity is long-lasting (i.e.,

there is no return to S from R); that everyone comes into contact with everyone else; that everyone is equally susceptible; and, that the population size doesn't change [58]. Over short time periods and in small communities these assumptions may be reasonable, yet in other contexts, a model making different assumptions may better capture the underlying process. As a result, many other disease models have been crafted since Kermack and McKendrick's initial publication. Approaches vary depending on the research question being addressed; some strive to accurately represent realistic contact patterns [73, 97] while others focus on capturing the possibility of superspreading, when some individuals infect an unusually large number of people [68].

1.2 Our Contribution

We seek to answer the question: How does the spread of a respiratory virus (we specifically consider SARS-CoV-2) depend on contact patterns and behavior? To do so, we build a coupled disease-behavior dynamical model that emphasizes the course of infection for COVID-19 cases, heterogeneous contact patterns, and shifting behaviors. Here we will provide a brief background on previous work in these aspects of disease modeling to situate our contribution. In Chapter 2, we discuss in more detail the expansion of the compartments of our model beyond S, I, and R to better reflect the stages of COVID-19 infection. Our approach to incorporating heterogeneous mixing and rational decision-making using network influence games is covered in Chapters 3 and 4, respectively. By simulating the spread of SARS-CoV-2 in different scenarios, we can weigh how different variables, such as resistance to masking, perceived risk of infection, and social influence, interact and affect patterns of disease transmission (Chapter 5). The implications of our findings are discussed in Chapter 6.

1.2.1 Progression of Disease

COVID-19 infection is more accurately described by using additional epidemiological compartments beyond S, I, and R. The SIR progression assumes that individuals become infectious immediately upon infection and that all individuals are equally infectious. For many diseases, including measles and COVID-19, this simplification obscures key characteristics of the course of infection [95, 10]. Ergo, modelers may instead use an $SEIR$ model. Here the E compartment represents individuals who are Exposed to the disease but are not yet infectious [95]. The Exposed compartment serves as a time delay which can be relevant when analyzing disease dynamics. Furthermore, individuals infected with COVID-19 can be infectious before showing symptoms or never show symptoms at all [108]. This asymptomatic infectious state can be differentiated from symptomatic infectious individuals by using an $SEAIR$ model where A represents Asymptomatic infectious individuals [100, 101, 12]. We will use a progression similar to $SEAIR$ in our model to account for these two key features of COVID-19 infection that aren't captured by the SIR progression: delay between infection and infectiousness, and asymptomatic infectious individuals (see Figure 2.2). Further details about this part of the model can be found in Chapter 2.

1.2.2 Heterogeneous Contact Patterns

Human contact patterns are more variable than the well-mixed SIR model suggests and this variability affects disease dynamics [104, 9, 75]. For example, humans do not mix equally with people of all ages; most daily interpersonal contact occurs between people of the same age group [79, 64]. Modelers can account for this variability in contact patterns in different ways. One technique in the ODE framework is to create additional compartments within the epidemiological classes S, I, and R, such that individuals of the same age mix with each other more than individuals of other age groups [33, 113, 70]. This approach can be generalized by having different rates of local (within neighborhood) and global (outside neighborhood)

mixing [8]. Neighborhoods may be broadly defined so that the groupings can be based on a variety of characteristics, including age and/or spatial location [8].

An alternative approach to describe heterogeneous contact patterns is to use a network to represent potential disease-spreading interactions [84]. Networks are a collection of nodes denoting individuals and edges (links between nodes) that denote interactions between individuals. Depending on the research question, modelers may generate random networks (more in Chapter 3) or build their own network structure to emphasize a particular social pattern [77]. Social patterns of emphasis may include household contacts or trips to common locations, e.g. supermarkets or hospitals [71, 92]. Networks have been used to model many diseases, including influenza [69, 62], smallpox [71, 88], and Ebola [93], as well as COVID-19 [89]. While networks require more computational power than ODEs and are not always analytically tractable (thus requiring simulation), they can often capture complex and varying contact patterns. We take advantage of these properties in Chapter 3 when building our own network for COVID-19 transmission.

1.2.3 Shifting Behaviors

Finally, we would like our model to account for shifting human behaviors, such as masking and social distancing, which can change individual susceptibility to infection and modify typical contact patterns. Susceptibility is naturally thought to vary throughout populations [65, 28], but as individuals alter their behavior to prevent infection during an outbreak, even greater differences in susceptibility may arise [36]. Some people will take precautions, reducing their susceptibility, and others will not. Likewise, these differing behavioral choices will alter the typical contact patterns outlined above as some people refrain from spending time with each other [60, 37]. Consequently, these behaviors can substantially impact disease dynamics and are important to account for in disease models.

There are several ways to incorporate behavior dynamics into disease models and we summarize a few of those approaches here. We call these types of models coupled disease-

behavior dynamical models because the disease and behavior dynamics depend on and are influenced by one another [106]. Coupled disease-behavior dynamical models can be applied to both homogeneously and heterogeneously mixed populations. Because we will be using a network to model interactions, we will focus here on how disease-behavior dynamics have previously been studied on networks. Wang and colleagues summarize a variety of implementations but ultimately sort these models into four categories based on the type of network: lattice or static networks, multilayer networks, adaptive or time-varying networks, and empirically-derived networks [106]. We will focus on static and multilayer networks.

An early coupled disease-behavior dynamical model was designed by Funk and colleagues in 2009 [41]. Using static networks, the authors proposed that disease awareness would spread much like disease, except the quality of awareness information would decrease as it got farther from the source (an infected individual) [41]. Individual disease susceptibility was linked inversely with the quality of awareness information [41]. To assume that awareness spreads in the same interactions as disease infection is perhaps unrealistic, so Funk and colleagues went a step further by using two potentially different networks, one for disease transmission and another for awareness propagation. This combination of networks is typically called multilayer or multiplex [61, 42, 43]. These networks have multiple layers composed of the same set of nodes but different sets of edges. One set of edges reflects potential infection-causing interactions, while another represents potential information-exchanging interactions [15]. This separated representation of disease and behavior spread allows the two entities to spread at different rates and incorporates the reality that information can be exchanged using current technology without engaging in an interaction where transmission is possible.

Funk and colleagues state that awareness may cause individuals to change their behavior but associate no cost with this shift [41]. In reality, if awareness means taking preventive measures to avoid infection, these actions would likely incur a cost. This cost may be social (e.g. pressure from friends to get together rather than stay home), political (e.g. only Democrats wear masks type sentiment), physical (e.g. masking worsens asthma), or

economic (e.g. can't work from home and have only one source of income). The relative cost of taking precautions compared to the risk of infection if one doesn't change one's behavior is a calculation many individuals have to make.

Often, in well-mixed models, this type of calculation can be described using game theory, a concept introduced by economists to model rational strategic decision-making [105]. A game is composed of two or more players with two or more possible actions. Each set of joint actions has a set of associated payoffs for each player which summarize the cost or benefit of that joint scenario to each individual. Based on this known set of payoffs, players will simultaneously decide which action to take to maximize their payoff (benefit). For example, Reluga uses a game-theoretic construct to model the adoption of social distancing behavior in a well-mixed population during an epidemic [90]. In this implementation, social distancing has an associated cost and is adopted if infection risk is sufficiently high [90]. Eliminating the homogeneous-mixing constraint and applying this game-theoretic approach to a networked population is desirable but also more complicated.

One technique for applying game theory on networks is to use network influence games [55]. In these influence games, each node (individual) has a threshold level and weighted connections with other nodes. The node can perform one of two actions based on how it is influenced by its neighbors and whether the neighbors' influences are strong enough to exceed the node's threshold for changing its action. As far as we know, this approach has not been used in coupled disease-behavior dynamical models to describe individual behavior choices and social influence. As we discuss further in Chapter 4, this approach will allow us to incorporate individual resistances to masking, perceived risks of infection, and social pressure from neighbors into the decision-making process about behavior change during an epidemic.

Chapter 2

Disease Progression

As we discussed in Chapter 1, infectious disease models can critically inform policy-making before and during an epidemic. Here, we expand on the implementation of early infectious disease models and map the disease compartments of our model to those used in previous work.

2.1 Compartmental Models

Recall from Chapter 1 the classic SIR compartmental model introduced by Kermack and McKendrick in 1927 [58]. Individuals begin in the susceptible state. Once infected they move to the infectious state, and finally to the recovered or removed state when they are no longer infectious (Figure 2.1). Typically, the model is seeded with some number or proportion of initially infectious people so that there is potential for an outbreak to occur. The system is closed (it does not account for births or deaths) so $S + I + R = N$ where N is the total number of individuals in the population. In its original implementation [58], this model was

parameterized using the following ordinary differential equations:

$$dS/dt = -\beta SI$$
$$dI/dt = \beta SI - \gamma I$$
$$dR/dt = \gamma I.$$

Here β is the probability of disease transmission *given contact* and γ is the rate of recovery, where $\frac{1}{\gamma}$ is the time it takes to recover.

There are three key assumptions made in the SIR model that we reconsider in our coupled disease-behavior dynamical model for COVID-19. First, the SIR implementation limits the possible epidemiological states of disease. In this chapter, we further explore this assumption and how our model grapples with it. Second, the SIR ODE model assumes homogeneous mixing among individuals. Homogeneous mixing is captured in the SI product term as part of both the dS/dt and dI/dt equations where all S and I individuals can come into contact with each other. We address this assumption using contact networks in Chapter 3. Finally, the SIR model does not allow for varied susceptibility or behavior changes among individuals, as all the individuals in S are seen as identical. We tackle this assumption using influence games in Chapter 4.

Below we present some illustrative and common examples of compartmental disease models to inform our work on COVID-19 disease progression.

Figure 2.1: Movement of individuals through SIR compartments.

The differential equations describing the SIR model limit the epidemiological disease states to three categories, excluding the possibility of reinfection or different clinical manifestations of disease. This choice of compartments is suitable for a disease in which there is long-lasting immunity – longer at least than the time scale over which the model is being run. For example, mumps and pertussis can both be modeled using the three SIR compartments [17, 94]. Some diseases, however, do not have long-lasting immunity. Influenza is one example; if you make it through this year's flu season you will likely be immune until the following season, at which point you will be susceptible again to the circulating strain. If one is building an influenza model spanning multiple seasons, it makes sense to use an $SIRS$ model, where after a recovery period individuals move back into the susceptible state and the progression can repeat [52]. However, if there is no period of immunity, meaning that after a period of infectiousness individuals immediately return to the susceptible state, then an SIS model may be used. Gonorrhea is one disease to which SIS models have been applied [50]. For some diseases, there is no immunity or return to susceptibility but continuing infectiousness. HIV/AIDS fits this description and so may be modeled using an SI model [53].

In addition to the immunity assumption made by the SIR model, there is also the presumed instantaneous switch from susceptibility to infectiousness. An instantaneous switch is clearly not realistic; it takes time for a virus or bacterium to replicate sufficiently to get to levels where it is transmissible to others. The period of time from initial infection to onset of infectiousness is called the *latent period*. At times, modelers may assume the latent period is negligible, especially when it is short relative to the infectious period [4]. For some diseases, however, it may be critical to include the latent period in a model so that the timing of individual infectiousness is closer to reality, especially for diseases, like tuberculosis, where the latent period is long relative to the infectious period [85]. In these cases, an $SEIR$ model is often used [2]. The E compartment represents Exposed individuals who become infectious upon reaching the I compartment. Ebola is one disease to which an $SEIR$ model has been

applied [2]. Furthermore, just like the SIR model, the $SEIR$ model can be extended to represent the case when individuals return to a susceptible state after a period of recovery, resulting in an $SEIRS$ model [31].

Slightly different from the latent period is the *incubation period* of a disease, or the time from initial infection to symptom onset [24]. If the incubation period is longer than the latent period, then there is an interval of time when individuals are not showing symptoms but are infectious. If the individual will never develop symptoms, then they have an *asymptomatic* infection. If the individual is not yet showing symptoms but will eventually develop them, then we consider them to be in a *presymptomatic* state. Asymptomatic or presymptomatic periods can also be incorporated into compartmental disease models. For instance, measles has been modeled using an $SEAIR$ model [1]. In this approach the A compartment stands for Asymptomatic, but the individuals eventually move into a symptomatic infectious state, I, so by our definitions above we would consider A to be a presymptomatic compartment. Sometimes this system is denoted as an SEI_1I_2R model, where I_1 is a presymptomatic state and I_2 a symptomatic state; see [19] for an application to rabies. In the next section, we discuss one way to separate the trajectories of asymptomatic and symptomatic infectious individuals. There are many other possible compartmental progressions, and a single disease may be modeled using multiple different progressions depending on the research question and context [16].

2.2 Our Disease Model

To capture the progression of COVID-19 infection, we have created a compartmental disease model inspired by SEI_1I_2R models and expanded to include truly asymptomatic individuals. As shown in Figure 2.2, all individuals (except any seed infections) begin in the susceptible (S) state and progress to an exposed (E) state. The E compartment acts as a time delay between infection and the onset of infectiousness; data from China suggest that this delay is relevant to COVID-19 infection and transmission [67]. From the exposed state, individuals

can take one of two routes depending on whether or not their infection will ultimately be symptomatic or remain asymptomatic. Based on data from the CDC, as of March 2021 [25], we estimate that about 70% of infections become symptomatic, while 30% remain asymptomatic. Thus, 70% of individuals move to I_1 (presymptomatic) and on to I_2 (symptomatic), where they will exhibit symptoms. The other 30% of individuals progress through the disease via the A_1 and A_2 (asymptomatic) compartments. After some time, I_2 and A_2 individuals will move to the removed or recovered state where they are no longer infectious.

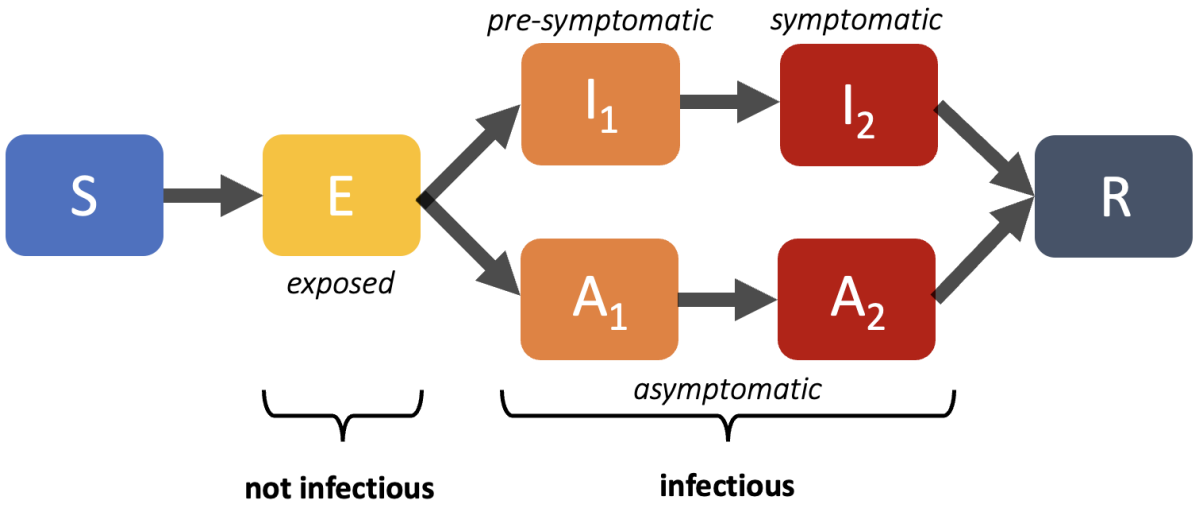

Figure 2.2: Progression of infection and disease via symptomatic or asymptomatic routes in our model. The E compartment is important for the time delay between infection and infectiousness. Distinction between infectiousness states I_1, I_2, A_1 and A_2 allows for differentiation of infectiousness in these stages, as indicated by preliminary data [49, 27, 7].

We make this distinction in infectiousness compartments for a couple of reasons. First, several studies [49, 27, 7] have indicated that individuals may be most infectious just before showing symptoms or at time of symptom onset. We create I_1 to allow for this possibility and differentiate I_1 from A_1 should the data indicate that asymptomatic infectiousness peaks at a different time. Although currently there is a limited understanding of how the relative infectiousness of asymptomatic individuals compares with that of symptomatic cases over time, this two-compartment asymptomatic implementation provides flexibility should further data become available. Second, individuals in I_1 may behave more like susceptible

or asymptomatic individuals if they do not know they are sick, compared to individuals in I_2 who are likely to feel unwell and reduce their contact with others. Thus, differentiating these compartments will allow us to consider differences in behavior.

Several biological factors control the movement of individuals through these compartments. Before introducing those components of the model, though, it is important to note that because we will be implementing this disease model on a network, we will be using discrete time steps. With that in mind, we can begin to consider how individuals move from one compartment to the next.

Each day, movement from S to E happens upon infection, so this process requires contact between infectious and susceptible individuals. The parameter β describes the baseline probability of transmission given contact between two individuals. As mentioned earlier, the division of infectious states into four compartments allows us to instantiate the model with different levels of transmissibility depending on stage and symptoms of infection (Figure 2.3). When we discuss our network structure in Chapter 3 and the possibility of taking prophylactic measures in Chapter 4, we will see how these different factors further modify the probability of transmission.

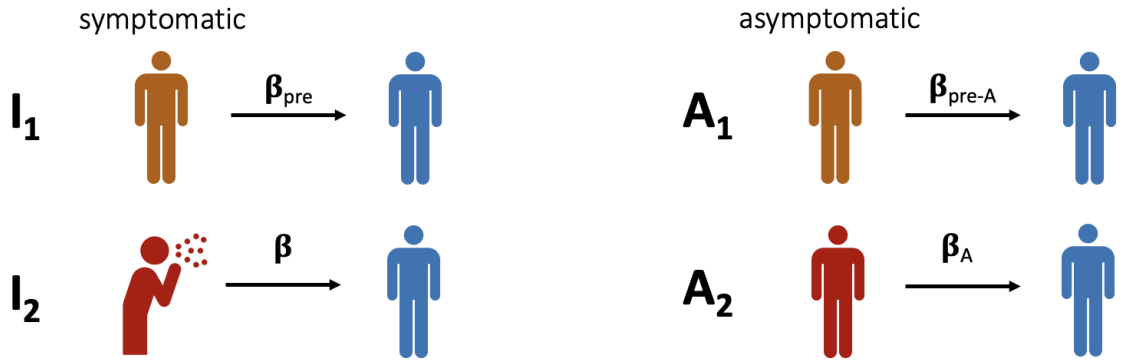

Figure 2.3: Varying probabilities of transmission given contact between infectious (orange and red) and susceptible (blue) individuals depending on symptoms and state of infectiousness of infectious individual.

Movement from E to I_1 or A_1 occurs after 2 time steps (Figure 2.4) [72, 67]. This time

is independent of whether the individual enters I_1 or A_1 since we have not seen evidence (to date) to suggest that the timing would differ between asymptomatic and presymptomatic individuals.

Individuals usually show symptoms 4 to 5 days after infection (Figure 2.4) [44, 66]. Thus, movement from I_1 to I_2 occurs after 3 time steps. Movement from A_1 to A_2 occurs the same way because we have not seen data suggesting otherwise.

Finally, movement from I_2 to R occurs after 5 time steps (Figure 2.4) [26, 49]. The same process is used to determine how long individuals are in the A_2 compartment, as we haven't observed data showing that these two disease states have different infectious period lengths.

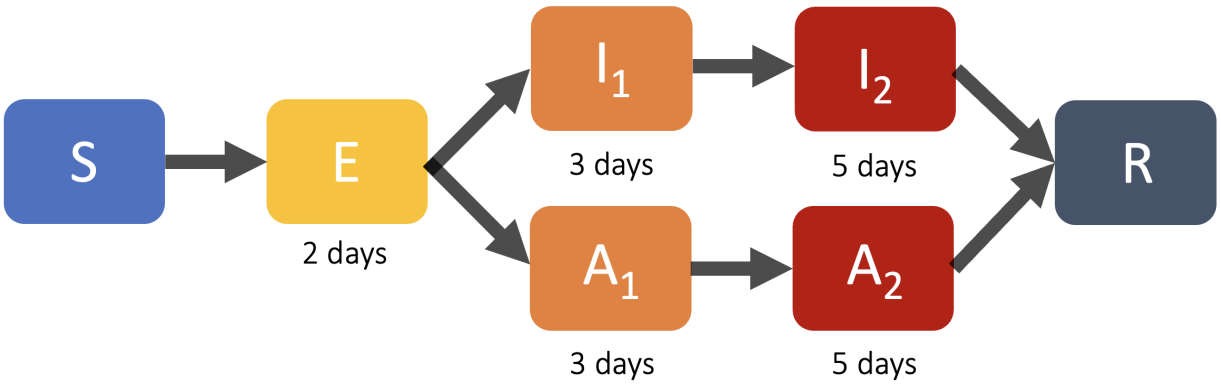

Figure 2.4: Number of time steps (days) that individuals spend in the exposed and infectious compartments.

In total, individuals are infectious for 8 time steps (days). This timing likely aligns with the period when cases would be most infectious, as the duration of infectiousness for COVID-19 varies substantially across individuals and infections [20, 49]. Because they are model parameters, these durations can easily be updated in the future to reflect the most up-to-date knowledge. Future work may also look to draw these values from probability distributions that reflect the variability in duration of infectiousness.

With the compartmental disease process in place, in the next section, we turn to the contact patterns through which individuals become infected.

Chapter 3

Networks and Contact Structures

Human contact patterns are heterogeneous in nature. This variation demonstrably impacts the spread of infectious diseases [104, 9, 75]. Depending on the aims of a modeling study, including the population of interest and route of disease transmission, accounting for this diversity in social contacts in the model can increase model accuracy [9]. In Chapter 1 we gave a brief overview of the array of approaches modelers take to include contact heterogeneity and established that we would be using contact networks to do so in our own model. Here, we introduce network terminology, three standard types of networks, and the network we build.

3.1 Networks Overview

Networks are defined as a collection of nodes and edges. *Nodes* represent individuals, or groups of individuals, and are connected by *edges* (Figure 3.1). These edges may be directed (i.e., have a start and endpoint) or undirected. If two nodes are connected by an edge we call these nodes *neighbors*. Similarly, a sequence of nodes such that there is an edge between any two consecutive nodes is called a *path*. Both nodes and edges can have assigned properties. One fundamental node property is its degree, defined by its number of edges. If a network is directed, then each node would have a separate in- and out-degree. An example of an edge

property is an edge weight, assigned when the network is constructed and often designed to represent the strength of a connection. We will see other node and edge properties when we construct our own network disease model later in this chapter.

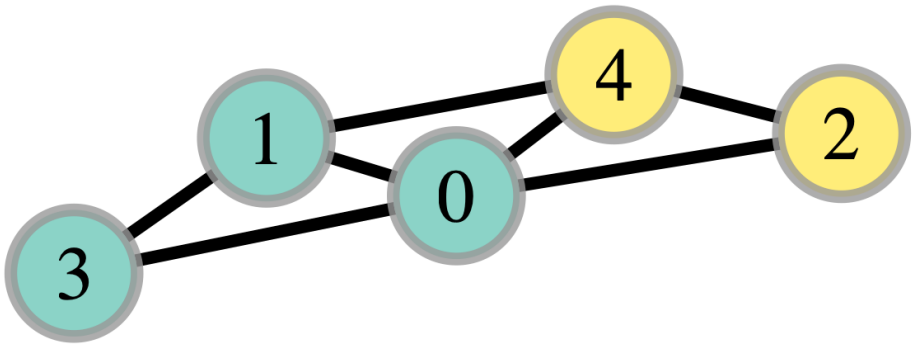

Figure 3.1: Example of an undirected five node network. The blue and yellow circles are nodes, and the black lines connecting them are edges. Node 0 has a degree of 4, while nodes 1 and 4 have a degree of 3, and nodes 2 and 3 have a degree of 2.

Networks can be used to represent relationships between individuals or groups. For instance, nodes might be scientists and edges may connect those who have co-authored a paper [83]. Additionally, edges could be used as more active paths for information or disease exchange. Namely, nodes might be cities and edges might be flights between those cities and the consequent movement of passengers or cargo; Colizza *et al.* [30] and Hufnagel *et al.* [54] both use these types of networks to model infectious diseases. Alternatively, in the context of HIV/AIDS, nodes might represent individuals and edges might connect those that have had sexual contact [46].

Real-world networks of human contact patterns have a few key properties. First, real-world networks are typically highly clustered [102]. Clustering can be thought of as the chance that two nodes who have a mutual friend are also friends. For instance, if nodes A and B are connected by an edge and nodes A and C are connected by an edge, what is the chance that nodes B and C are connected (Figure 3.2)? If this chance is high, then the network would be considered highly clustered, and if this chance is low, then the network has low clustering.

Second, real-world networks typically exhibit the small-world property [102]. This property

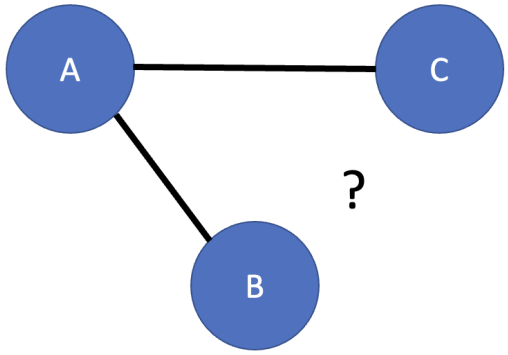

Figure 3.2: In highly clustered networks, nodes with mutual friends (e.g., B and C) are also likely to be connected. In networks with less clustering, nodes with mutual friends are less likely to also be friends.

states that any node can be reached from any other node by traversing relatively few edges (i.e., via a relatively short path). There is no strict definition of "relatively few edges," but the number six, as in "six degrees of separation," has appeared in popular culture [45] as well as sociology experiments [76]. A third, more controversial property of real-world networks is a power law (or scale-free) degree distribution (see Figure 5.30 in Chapter 5) [18]. This degree distribution is characterized by most nodes having a low degree, with a small number having a very high degree. Let $\mathbb{P}(k)$ denote the probability of a node having degree k. Then the power law degree distribution is defined by $\mathbb{P}(k) \sim k^{-\gamma}$, where k denotes degree, $\gamma > 1$, and a proportionality factor is required to normalize the distribution [11]. Efforts to generate synthetic real-world networks will typically try to include these properties.

There are several standard network construction approaches that vary in how well they represent real-world interactions; we review three relevant approaches here. One of the simplest networks to construct is an Erdos-Renyi random graph. In this type of network, each node is connected to every other node with some probability p between 0 and 1. The resulting degree distribution follows a Poisson distribution: $\mathbb{P}(k) = \frac{(Np)^k e^{-Np}}{k!}$ where k denotes degree and N is the total number of nodes. Erdos-Renyi networks have very little clustering

but exhibit the small-world property. They do not accurately represent large-scale daily human contact patterns, but they are easy to construct.

A second type of network is generated by the Watts-Strogatz model. This model begins with nodes arranged in a circle and connected to their k (an even integer) closest neighbors ($\frac{k}{2}$ neighbors each on the right and left). Human connections are not solely spatial, though, so Watts and Strogatz take an additional step to create a more realistic network. They use a rewiring process, whereby for each of its k edges, with a given probability β, a node can "choose" to remove that edge and form a new edge with another node. The initial connection only to a node's neighbors leads to high clustering, while the random rewiring allows these types of networks to satisfy the small-world property. Where the Watts-Strogatz model falls short of constructing a real-world network is its degree distribution; if the rewiring probability $\beta = 1$, then the degree distribution follows a Poisson distribution because the graph is identical to an Erdos-Renyi random graph. If, instead, $0 < \beta < 1$, then the degree distribution is approximately Poisson but has a sharp peak at degree k. This distribution does not have the heavy tail of high-degree nodes characteristic of real-world networks.

A final network construction method of relevance is the Barabasi-Albert preferential attachment model. This model adds nodes and edges in sequence – one node is added at each time step. When a new node is added to the network it forms m edges, where m is the sole model parameter. A newly added node forms an edge with an existing node i with a probability proportional to i's degree at that time. This probability is higher for nodes with more edges than nodes with fewer edges. Thus, these higher degree nodes will be more likely to grow their degree even further – hence the term preferential attachment. The degree distribution of Barabasi-Albert networks follows a power law distribution of the form $\mathbb{P}(k) = \frac{2m^2}{k^3}$.

In addition to these generic networks, scientists often construct their own networks specific to the population or system that they are studying. There are numerous ways to do this, as we touched on in Chapter 1. One common technique for disease models is to use

multilayer networks (see Figure 3.3) where each layer represents a different type or level of contact; e.g., household, school, work, social, and community [80, 77]. Depending on the implementation, the same nodes may exist in each layer but have a different set of contacts based on the context considered in that layer. For example, in [80], there are no adults in the school layer and there are no children in the adult layer, but all individuals are present in the household, social, and community layers. Layers may also vary in their degree distributions, levels of clustering, and small-world property adherence. Additionally, edges in each layer can be weighted differently to reflect the higher or lower chances of transmission in that context. While this type of network adds complexity, it allows for a more accurate representation of the variety of contacts individuals have daily.

3.2 Our Network

We create a realistic set of contact patterns using a multilayered network. The network consists of three layers: household, work/school, and friends. We make the simplifying assumption that workplace interactions are analogous to those in school. As a result, we do not incorporate age into the model. All three layers consist of the same nodes but have a different edge generation process and different edge weights, signifying the frequency of that type of contact (Figure 3.3). We make this distinction in edge weights because interactions in these three arenas carry different probabilities of disease transmission. Later, in Chapter 4, we will add another set of edge weights related to social influence, but our discussion, for now, is limited to the edge weights that, because they denote the frequency and duration of contact, directly affect the probability of disease transmission.

At the household level, we form many separate but completely connected groups of individuals. The complete connectedness of these groups is meant to represent that all individuals within the same household are likely to have close contact with one another. These group sizes are drawn from a discrete probability distribution generated from US American Community Survey Census data from 2018 (Figure 3.4) [99]. Households can

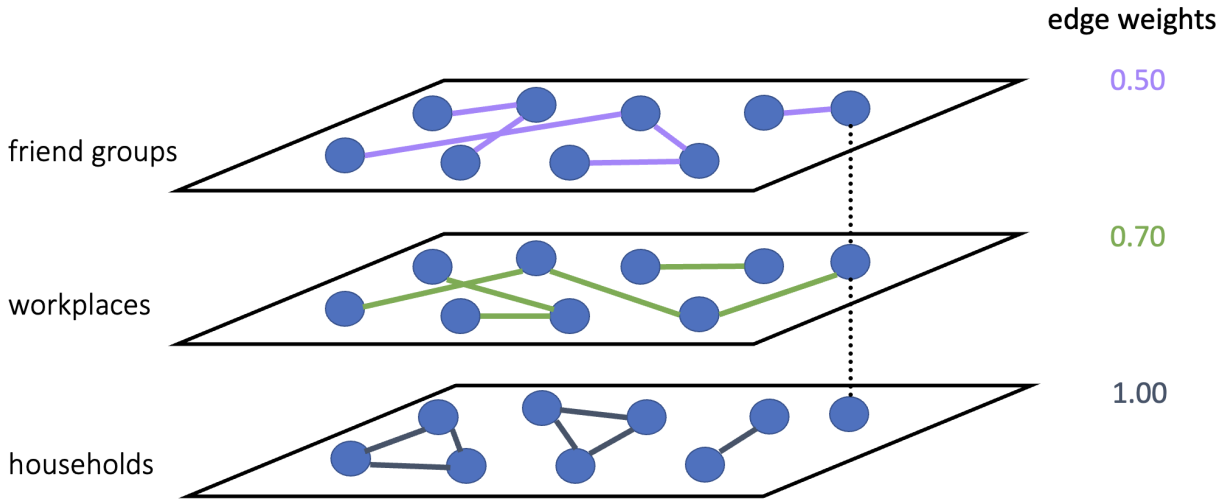

Figure 3.3: Multilayer network, with three layers: household, workplace, friend group. Each layer contains the same nodes, but the connections differ between layers and are assigned different edge weights depending on the layer.

range in size from 1 to 7 individuals, with a median of 3 individuals. Edges formed in the household layer carry a disease weight of 1 because the chances of disease transmission in household settings are high [74].

At the work level, we draw workplace sizes from a normal distribution centered at 20 with a standard deviation of 5. We then randomly assign nodes to workplaces. Within each workplace, we form edges using an Erdos-Renyi process with $p = 0.35$. These edges carry a disease weight of 0.7, or roughly 5/7, signifying the 5 out of 7 days a week that these individuals may come into close enough contact for disease transmission.

At the friend level, we draw friend group sizes from a normal distribution with mean 6 and a standard deviation of 2. We then randomly assign nodes to friend groups. Within each friend group, we form edges using an Erdos-Renyi process with $p = 0.7$. These edges carry a disease weight of 0.5, estimated based on less frequent but likely longer and closer contact than with work colleagues.

These values and distributions were chosen so that the degree distribution would be centered around 14, a realistic estimate for daily contacts, based on work by Mossong *et al.*

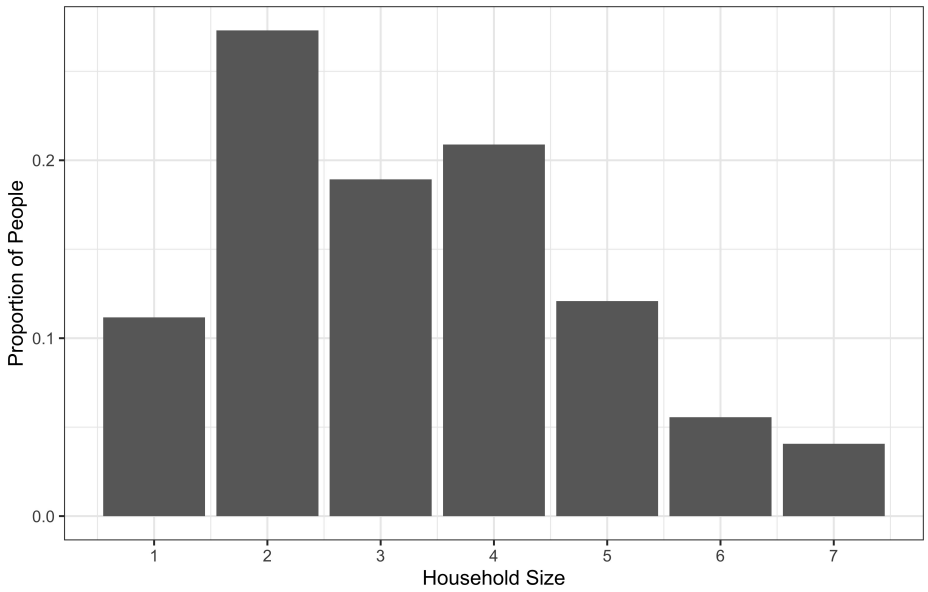

Figure 3.4: Distribution of household sizes based on US ACS Census data from 2018. All households with seven or more people are considered to have exactly seven people.

[79]. The workplace size distribution was loosely based on data from 2004 [59]. In the absence of data on friend group sizes, we chose a distribution we thought was reasonable. The means and standard deviations of these distributions, as well as the probability of edge formation and disease transmission edge weights, are all model parameters. Thus, the sensitivity of simulation results to these values can be tested in the future and values can be adjusted should further data become available. Heavier tailed degree distributions should also be experimented with in the work and friend group layers as an added element of complexity.

It is important to note that in our implementation, a pair of nodes can only be connected once; even if two nodes are in the same household and workplace, the household interaction "overrides" the workplace interaction given the higher edge weight. Similarly, if two nodes are in the same household and friend group, the household interaction will override the friend group one. Finally, because the workplace contact edge weight is higher than the friend group contact edge weight, a workplace interaction overrides a friend group one. This assumption simplifies the model and allows us to more accurately choose parameters that align with the transmissibility of SARS-CoV-2.

Networks are constructed using graph-tool version 2.35 [86].

3.3 Disease Spread on Networks

Given this network, we will simulate the spread of SARS-CoV-2. We seed the outbreak with a randomly selected node who is symptomatic infectious (I_2), up to three of their neighbors who are presymptomatic infectious (I_1), and up to four of their neighbors who are exposed (E) but will eventually become symptomatic. We seed each outbreak with the same maximum number of infected people so that we can compare our results. As the virus spreads, we keep track of the disease state of every node (S, E, I_1, I_2, A_1, A_2, or R) and the length of time that they have been in that disease state so that they can progress to the next state at the appropriate time step (discussed in Chapter 2). Disease transmission is possible across edges between infectious (I_1, I_2, A_1, A_2) and susceptible (S) individuals. The daily probability of transmission between a susceptible and infectious individual is given by $\beta \cdot e_{ij}$, where e_{ij} is the symmetric edge weight between the two nodes. We will modify this probability in the next chapter when we incorporate masking behavior.

Chapter 4

Behavior and Influence

4.1 Disease-Behavior Models

Many factors affect population disease dynamics. Contact patterns, as we reviewed in Chapter 3, are one such factor and varying behaviors are another. Behavioral effects may be a result of individual changes in the number or types (e.g., household, work, or friend) of contacts, in addition to increased or decreased chances of transmission when encountering contacts. Consequently, it can be important to incorporate these behavioral aspects and how they shift over time into a mechanistic disease model. Behavioral changes have been built into disease models in a variety of ways that depend on whether the model assumes homogeneous or heterogeneous contact patterns and how much individual preferences are taken into account [106].

In homogeneous-mixing models, behavioral changes are often implemented using either rule or economic-based methods [106]. Rule-based models use disease prevalence-based rules that lead agents to change their behavior under certain conditions. Economic-based models assume that individuals can calculate and attempt to maximize their personal well-being by making behavioral changes [106]. Poletti and colleagues built a rule-based homogeneous-mixing SIR model that has individuals alter their behavior when disease prevalence exceeds

a certain level where the risk of infection is too high to justify normal behavior [87]. In contrast, Chen and colleagues used an economic-based homogeneous-mixing SIS model where individuals determine how much time to spend outside their home interacting with others based on the prevalence of a disease and the cost of staying home [29]. Economic-based models may also employ game-theoretic techniques which we discuss in further detail in section 4.2.

In heterogeneous-mixing models, the implementation of behavior dynamics takes several forms depending on how the varying contact patterns are captured [106]. In Chapter 1, we introduced the implementation by Funk and colleagues as an early example of how disease-behavior dynamics may be modeled on networks [41]. As you will recall, the paper proposed that awareness spreads outwards from an infected individual with decreasing quality as it gets farther from the source. With awareness comes decreased disease susceptibility based on the assumption that aware individuals will take preventive measures [41]. Funk and colleagues employ multilayer networks, where disease transmission and awareness propagation layers of the multilayer network contain the same nodes but different connections among the nodes. Because interactions occur that cannot cause disease transmission but can affect how humans behave with respect to a disease, distinguishing between these two types of exchange is important. We pointed out in Chapter 1, however, that Funk and colleagues' approach assigns no penalty or cost for taking precautions. In reality, reducing one's contacts, wearing a mask, or getting a vaccination, for example, all incur some cost to the individual. While these costs may be small relative to the benefit – a reduced chance of infection – each individual must make this determination themselves. In our approach, detailed at the end of this chapter, we introduce a resistance value for each node which reflects the varying cost of masking for every individual.

Zhang and colleagues address this cost assumption in their SIR model that allows individuals to behave in one of three ways: laissez-faire, self-protective, or vaccinated [112]. Both self-protective and vaccination behaviors incur a cost that is less than that of infection.

To represent consecutive seasonal outbreaks of a pathogen like influenza, Zhang *et al.* run multiple rounds of behavior changes and infection spread. Between rounds, individuals can decide to imitate their neighbors' behaviors in the following season. While these behavior updates bring this model seemingly closer to the real-world decision-making process, the updates only occur after the prior season and before the next season begins, rather than *while* the disease outbreak is occurring. In consequence, the behavior update process doesn't incorporate disease prevalence. In our model, we will incorporate the disease prevalence into the decision-making process.

We seek to combine and refine these approaches to create a model which allows behavior and disease dynamics to influence each other while accounting for the fact that the two processes may happen at different rates. Furthermore, we build in the concept that changing one's behavior to prevent the spread of a disease can be costly and is influenced by an individual's neighbors' actions. To do so, we utilize game theory and a network-specific game-theoretic model, influence games.

4.2 Game Theory

Game theory provides a mathematical approach to modeling the rational decision-making of individuals in situations where these decisions are interdependent. Games are composed of players, each of whom plays an "action" based on what they expect other players to do. Associated with each vector of joint actions is a payoff for each player. This payoff can incorporate the costs of different actions. Players will always look to maximize their payoffs – called playing their *best response*. When each player plays their best response, the system has reached a state called *Nash equilibrium* [81], where no player has any incentive to change their action. We will revisit this idea of equilibrium in our model in section 4.4.

In the meantime, an illustrative and famous example of a two person game is shown in Figure 4.1. In this example, called Prisoner's Dilemma, two suspects must decide whether to confess to a crime [21]. The best outcome (highest payoff, -1) for both occurs if both do not

confess. However, if suspect 1 considers that suspect 2 will not confess, then it is in suspect 1's best interest to confess. Moreover, if suspect 1 considers that suspect 2 will confess, then it is still in suspect 1's best interest to confess. Through this set of calculations, under both scenarios, suspect 1 will decide to confess. The same calculation applies for suspect 2 who will also decide to confess. These best responses lead to a Nash equilibrium of (confess, confess).

		Suspect 2	
		Confess	Don't Confess
Suspect 1	Confess	-5, -5	0, -10
	Don't Confess	-10, 0	-1, -1

Figure 4.1: Prisoner's Dilemma two-player game theory example. Suspect 1 and 2 must each decide whether or not to confess to a crime they committed. The payoffs – the prison sentences – are given as Suspect 1, Suspect 2. The Nash equilibrium outcome, where both players simultaneously play their best response and have no incentive to deviate, is shown in blue.

Game theory has been used frequently in coupled disease-behavior models for vaccination decision-making [14, 13, 91]. When getting a vaccine, individuals weigh several potential costs, including the cost of infection, how expensive the vaccine is, how long it takes to get the vaccine, and potential side effects, among others. Depending on their personal preferences and the risks associated with the vaccine, they will decide whether or not an investment in vaccination is advantageous for them. Interestingly, high levels of vaccination can introduce a "free rider" effect where vaccination risks appear substantially higher than disease risks due to the nature of herd immunity [13]. This perception may cause many people to refrain from vaccinating. These free riders may not suffer the cost of becoming infected because

a sufficient number of other people are vaccinated, protecting them. However, if too many people try to take advantage of the free rider effect and choose not to vaccinate, then the disease can continue to spread, undermining their choice of not vaccinating.

It is important to note, however, that vaccination as a behavior is somewhat different from compliance with the other non-pharmaceutical interventions suggested or mandated to reduce the spread of COVID-19. Once one gets a vaccine, one can't "undo" it. Many of the behavior models we discuss below are applied to diseases like influenza, where vaccination is needed every year. These models allow for a yearly decision-making process after the course of an epidemic, as in Zhang *et al.* [112]. However, decisions to wear masks or social distance are happening on a different time scale. Individuals can change their decision about masking multiple times over the course of an outbreak, whereas a vaccination decision can only be made once during an outbreak. Thus, modeling the types of behaviors that can be changed more frequently requires a different approach.

A variety of approaches are used to capture the human decision-making process in vaccination models. Two common approaches include imitation dynamics [39] and evolutionary game theory. We briefly review a couple of models using these methods to demonstrate how they have been used in the context of vaccination and may be applied to non-pharmaceutical interventions, such as masking and social distancing.

Ndeffo *et al.* combine both imitation dynamics and payoff maximization to model vaccination decisions [82]. In their model, some fraction of individuals in a network update their vaccination behaviors by imitating an immediate neighbor, and the remaining fraction update by maximizing their perceived benefits. In the imitation process, a node i considers the behavior of one of its randomly selected immediate neighbors j. If i and j share the same behavior, then they stick with that choice. If their behaviors differ, then i adopts j's behavior with a certain probability based on if and how much higher j's payoff is. In Ndeffo's model, individuals can also update behavior to maximize their individual payoff in a process that weighs the cost of vaccination and infection, as well as the perceived probability of

infection based on their immediate neighbors' vaccination decisions. The update process is continually run until a steady state is reached (i.e., individuals' decisions no longer change over time), at which point vaccination occurs.

In the context of our goals with this project, an overall strength of this methodology is how it allows decisions to fluctuate and eventually converge to an equilibrium. A weakness is that imitation behavior is considered mutually exclusive to payoff-driven behavior. In other words, in Ndeffo's model, an individual does not take into account both the social influence of their neighbors and what their neighbors' decisions mean for their chances of infection. Furthermore, node i considers only the actions of a single neighbor in their decision-making process, rather than all or most of their neighbors. Imitating only a single random neighbor, instead of taking into account the behaviors of multiple neighbors, doesn't realistically represent the human decision-making process. Fukuda and colleagues [40] adjust this assumption by modifying the probability that i imitates j. In Fukuda's model node i still picks a random neighbor j, but instead of looking at only j's payoff, i calculates the average payoff of any individual adopting the same behavior as j. The average payoff is then used to determine whether to adopt the same behavior as j. This implementation reflects that a node i is likely basing their behavior decision off of the behavior of more than one other individual.

An alternative process is used by Xia and colleagues [111] to better capture social influence. Individuals use a combination of game-theoretical cost minimization calculations and the influence of neighbors to determine their vaccination decision. The cost minimization calculation is similar to those discussed earlier and involves the weighing of infection and vaccination costs. The social influence computation includes the strength of a connection, the behavior of a neighbor, and the number of neighbors accepting and rejecting vaccination. Individuals are then assigned a probability p that indicates their tendency to adopt the cost-minimizing behavior. With probability $1-p$ they instead adopt the behavior suggested by the social influence calculations. If both calculations point to the same decision, then this probability is irrelevant. Xia and colleagues' approach is closest to the one we consider

in our model of human behavior. Like Xia *et al.*, our decision-making model weighs both social influence and cost minimization. We use a construct called influence games, discussed in the next section, to calculate social influence.

4.3 Influence Games

Before we introduce influence games, it will be helpful to understand the *linear threshold model* [57]. The linear threshold model is defined as follows. In a network, a node i is influenced by each neighbor j with a weight $w_{ji} \geq 0$ such that $\sum_j w_{ji} \leq 1$. Each node i also has a threshold, or resistance, $q_i \in [0, 1]$. This threshold defines the weighted fraction of i's neighbors which must be "active" for i to choose to become "active". Otherwise, i will remain "inactive". Initially, all but a small set of nodes are inactive. Then the diffusion process begins: the active nodes can influence their neighbors to adopt an active state, who can influence their neighbors to adopt an active state, and so on. The diffusion proceeds until either all nodes are active or nodes are no longer changing their behavior. This idea of activity could be applied in many contexts, e.g. for i to choose to be vaccinated or to choose to mask. In the linear threshold model, due to non-negative influence weights, if a node becomes active at any point it remains in that state throughout the duration of the model. When thinking about vaccination, this continued activation may be acceptable, but in the context of behaviors like masking which can be "undone", we would like a model that allows for more switching of behaviors.

In 2014, Irfan and Ortiz [55] introduced linear influence games as a game-theoretic method to model the behavior of a finite-networked population. They focus specifically on stable outcomes. Similar to the linear threshold model [57], this approach begins with a directed network composed of nodes with varying threshold, or resistance, levels, $q_i \in \mathbb{R}$. Edges carry influence weights, $w_{ji} \in \mathbb{R}$ which reflect the amount of influence that a node j exerts on a neighbor i. These directed edges are not necessarily symmetric: a node i may exert more influence on a node j than j exerts on i. Each node in the network adopts one of

two behaviors, denoted by the integers -1 and 1. If there is no influence, a node with a positive threshold will adopt behavior -1 while a node with a negative threshold will adopt behavior $+1$ (see equation 4.1 below). A node i determines their incoming influence from all neighbors j using the following expression:

$$\sum_{j \neq i} x_j \cdot w_{ji}.$$

Comparing the value of this expression with their individual threshold q_i, a node will decide whether or not to change their behavior to maximize their payoff:

$$i\text{'s payoff} = x_i\Big(\sum_{j \neq i} x_j \cdot w_{ji} - q_i\Big). \tag{4.1}$$

This payoff calculation takes into account the actions of other players.

Equation 4.1 determines how i chooses whether their behavior x_i is -1 or 1; this is i's best response. If

$$\sum_{j \neq i} x_j w_{ji} - q_i < 0, \tag{4.2}$$

then it is in i's best interest to perform behavior $x_i = -1$ so that

$$x_i\Big(\sum_{j \neq i} x_j w_{ji} - q_i\Big) > 0,$$

maximizing i's payoff. One scenario in which equation 4.2 holds is when the total influence on i is 0 and i has a positive threshold. Alternatively, if

$$\sum_{j \neq i} x_j w_{ji} - q_i > 0, \tag{4.3}$$

then i chooses $x_i = 1$ so that i's payoff is largest. One scenario in which equation 4.3 holds

is if there is no net influence on i and i has a negative threshold. In the case where

$$\sum_{j \neq i} x_j w_{ji} - q_i = 0,$$

i will choose either action with equal probability. At each time step, each node sums the incoming influences to determine the action to perform at the next time step to maximize their payoff. Unlike the linear threshold model, this influence game model allows nodes to switch their behaviors back and forth. It is with this model in mind that we detail how individuals decide which preventive behaviors to take in our implementation.

4.4 Our Behavior Model

Given the disease progression and contact structure we established in Chapters 2 and 3, we can now construct the behavioral model. We focus specifically on masking behavior. Using an influence game approach, we will consider three components of individuals' decision process about masking: resistance, risk, and influence.

Individuals naturally have some level of resistance to masking. The monetary or temporal cost of finding or buying masks and the physical discomfort or political consequences of wearing a mask likely contribute to this resistance. An individual's baseline willingness to mask is then represented by $\tilde{q}_i \in [0, L]$, where L indicates a strong unwillingness to mask and 0 indicates no resistance to masking. We experiment with the scaling factor L in Chapter 5 to find an appropriate range of resistance relative to the remaining components of the behavior calculation. At model initialization, each individual, i, will be assigned $\tilde{q}_i$ drawn from a uniform distribution $\mathcal{U}(0, L)$. One consequence of this resistance structure is that individuals will choose not to mask in the absence of disease and social influence.

The second component of an individual's decision calculation is their perceived risk of

infection. One way to quantify risk is:

$$\text{risk of an event} = \text{cost of that event} \cdot \text{probability of that event.}$$

To calculate the perceived risk of infection for node i, ρ_i, we take the product of the perceived cost of infection, c_i, and the perceived probability of infection. We reason that an individual's perceived probability of infection is the proportion of their neighbors who are infectious. This construction assumes some level of baseline testing so that a node knows how many of their neighbors are infectious, even if presymptomatic or asymptomatic. Let U_i represent the set of node i's neighbors. Then $(A_1 \cup A_2 \cup I_1 \cup I_2) \cap U_i$ gives the set of all infectious neighbors of i and the perceived risk of infection is given by:

$$\rho_i = \sqrt{c_i \cdot \frac{|(A_1 \cup A_2 \cup I_1 \cup I_2) \cap U_i|}{|U_i|}}$$

where $|.|$ denotes cardinality. Infection costs, c_i, are drawn from a uniform distribution $\mathcal{U}(0, 1)$ and assigned to each individual at model initialization. These costs are meant to represent how important it is to an individual that they do not become infected. For example, older and immunocompromised individuals may have larger perceived costs of COVID-19 infection, as may people aware of the possible long-hauler effects of COVID-19 infection [5, 22, 98].

Both terms in this risk calculation fall between 0 and 1, meaning that their product is sublinear. Social influence, which we discuss next, is constructed linearly. Thus, we take the square root of the product so that the perceived risk of infection is comparable to social influence.

Following the linear threshold model of Kempe, Kleinberg, and Tardos [57], as well as Irfan and Ortiz's influence game model [55], we define the social influence on each node i to be:

$$\sum_{j \in U_i} x_j w_{ji}, \tag{4.4}$$

where, as above, U_i is the neighborhood of node i. The behavior of node j is encoded

by $x_j \in \{-1, 1\}$ where -1 indicates not masking and 1 indicates masking. The weight, or influence, of node j's actions on node i is denoted by $w_{ji} \in [0, 1]$ with larger weights denoting stronger influence. The incoming weights, w_{ji}, to node i are determined as follows. First, preliminary weights, w'_{ji}, are drawn from $\mathcal{U}(0, 1)$. A maximum influence sum, ϕ_i, is also drawn from $\mathcal{U}(0, 1)$. The preliminary weights, w'_{ji}, are then normalized so that $\sum_{j \in U_i} w_{ji} = \phi_i$. This process ensures that the total social influence on node i, given by equation 4.4, lies in $[-\phi_i, \phi_i] \subset [-1, 1]$, and at the same time creates random heterogeneity in the power of social influence on individual decision-making across the network. In Chapter 5, we explore the impact of the strength of influence on coupled disease and behavior dynamics by reducing the domain of the uniform distribution from which ϕ_i is drawn.

Bringing these three decision components together, node i will make a decision for the following time step by evaluating

$$D_i = \sqrt{c_i \cdot \frac{|(A_1 \cup A_2 \cup I_1 \cup I_2) \cap U_i|}{|U_i|}} + \sum_{j \in U_i} x_j w_{ji} - \tilde{q}_i. \tag{4.5}$$

If $D_i > 0$, then i will adopt masking behavior at the next time step. If $D_i < 0$, then i will not adopt masking behavior at the next time step. If $D_i = 0$, then i chooses either behavior with equal probability.

Equation 4.5 is the behavioral model without equilibration. However, this implementation can lead nodes to frequently (and unrealistically) switch their masking behavior back and forth as a result of incoming influences. To address this issue, we allow the behavior process to equilibrate at each time step: nodes go through an iterative process of deciding their masking behavior based on the choices of their neighbors until they no longer want to change their decision. When the system has reached equilibrium, nodes are playing their best response, and the chosen behaviors are considered a Nash equilibrium at that time step. This equilibration process happens between rounds of disease spread on the network. Thus, behavioral decisions can change more frequently than infection can spread in our model, as

in real life [56]. In Chapter 5, we explore the impact of this equilibration assumption by comparing model output with full equilibration, partial equilibration, and no equilibration.

With our complete coupled disease-behavior dynamical model in place, we can now turn to simulating outbreaks and examining the effect of these different elements of decision-making on outbreak dynamics.

Chapter 5

Modeling Results

We can now explore the interplay of mechanisms governing the behavior and disease dynamics of SARS-CoV-2.

5.1 Parameters

We use several parameters in our simulations (Table 5.1). At this stage, in choosing these parameters, we look to create a reasonably realistic infection process from which to deduce qualitative insights about the interactions among the disease and behavior mechanisms. If this model were to be used to predict case counts or other COVID-19 statistics, parameter choice, sensitivity, and uncertainty quantification would be critically important.

Recall that β represents the probability of transmission given contact between an infectious and a susceptible individual. In our model, this baseline β neither accounts for the intensity of contact (as represented by household, work, or friend edge weights in our multilayer network) nor how masking of either individual may reduce the chances of transmission. In reality, β is very difficult to estimate. We come up with some plausible values listed in Table 5.1 based on current data. Researchers estimate the average number of secondary SARS-CoV-2 infections caused by an infectious individual in a fully susceptible population to be 2.5 [25]. On average, each infectious node in our network will have a degree around

13 or 14 and be infectious for 8 days. To infect 2.5 of their susceptible neighbors in this time period, the probability of transmission given contact must be around 0.024. Given that asymptomatic individuals are thought to be 75% as infectious as symptomatic individuals [25], we chose the β values below.

Table 5.1: Parameter list.

Parameter	Value
β_{pre-A} (for individuals in A_1)	0.02
β_{pre} (for individuals in I_1)	0.02
β_A (for individuals in A_2)	0.02
β (for individuals in I_2)	0.033
Probability of developing symptoms	0.7
Time in E	2 time steps
Time in I_1 or A_1	3 time steps
Time in I_2 or A_2	5 time steps
Reduced chance of transmission if infectious person masks	0.3
Reduced chance of transmission if susceptible person masks	0.8
Workplace size	Normal(20,5)
Workplace edge probability	0.35
Workplace disease edge weight	0.7
Friend group size	Normal(6,2)
Friend group edge probability	0.7
Friend disease edge weight	0.5
Resistance scaling factor[1], L	0.2

[1]Model results for different values of L are shown in Figures 5.9, 5.10, 5.11, and 5.12

5.2 Network Characteristics

We use several different networks in our simulations. First, we use the multilayer network structure outlined in Chapter 3. This contact structure differentiates between household, work, and friend interactions when running the infection process. In the following experiments, we use four different one thousand node multilayer networks, with the degree distribution of each shown in Figure 5.1. As expected given the Erdos-Renyi process used to connect individuals in the work and friend layers, the degree distribution looks Poisson, centered around a degree of 13 or 14. Future work could vary these distributions further so that the degree distribution has a heavier right tail, reflecting the few individuals who have an unusually high number of contacts. This exploration would also be important if we wanted to incorporate the possibility of superspreading into our model [68].

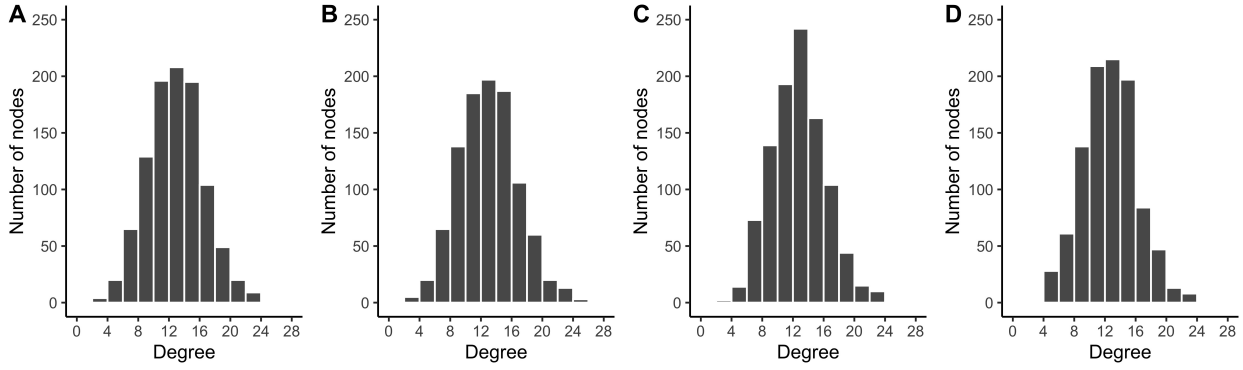

Figure 5.1: Degree distributions of four networks of one thousand nodes generated using the multilayer method outlined in Chapter 3. (A) network 1, (B) network 2, (C) network 3, (D) network 4.

Second, to compare different contact structures we generate two Erdos-Renyi random networks and two Barabasi-Albert preferential attachment networks of one thousand nodes each. You will recall that the degree distribution of Erdos-Renyi random networks approaches a Poisson distribution as the number of nodes approaches infinity, and we observe a similar degree distribution in our networks of 1000 nodes (Figure 5.2). The Barabasi-Albert preferential attachment networks have a power law degree distribution (Figure 5.3).

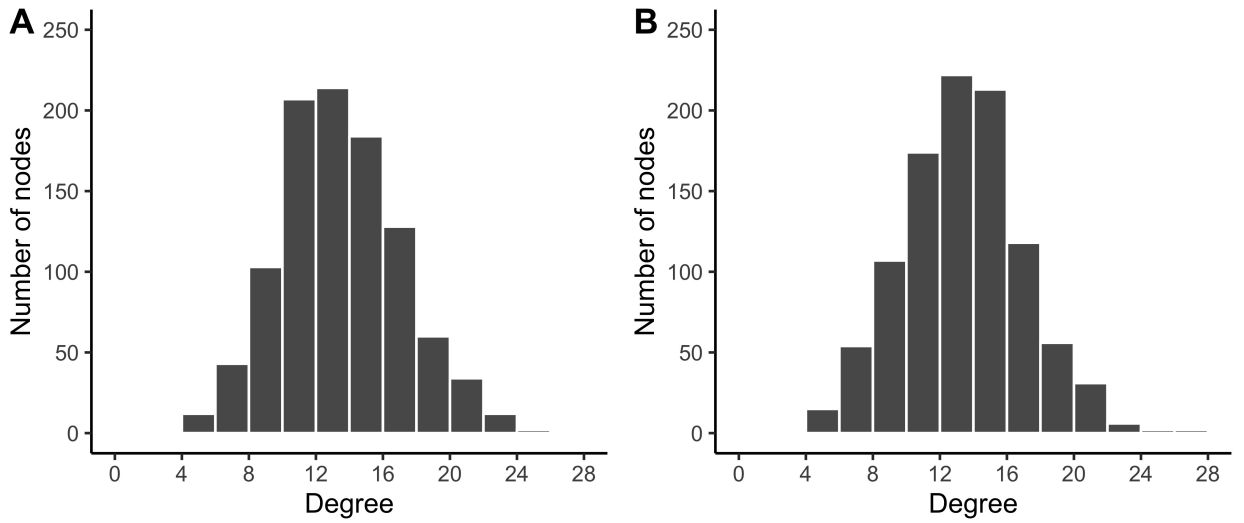

Figure 5.2: Degree distribution of Erdos-Renyi random networks of one thousand nodes generated with $p = 0.014$. (A) Erdos-Renyi network 1, (B) Erdos-Renyi network 2.

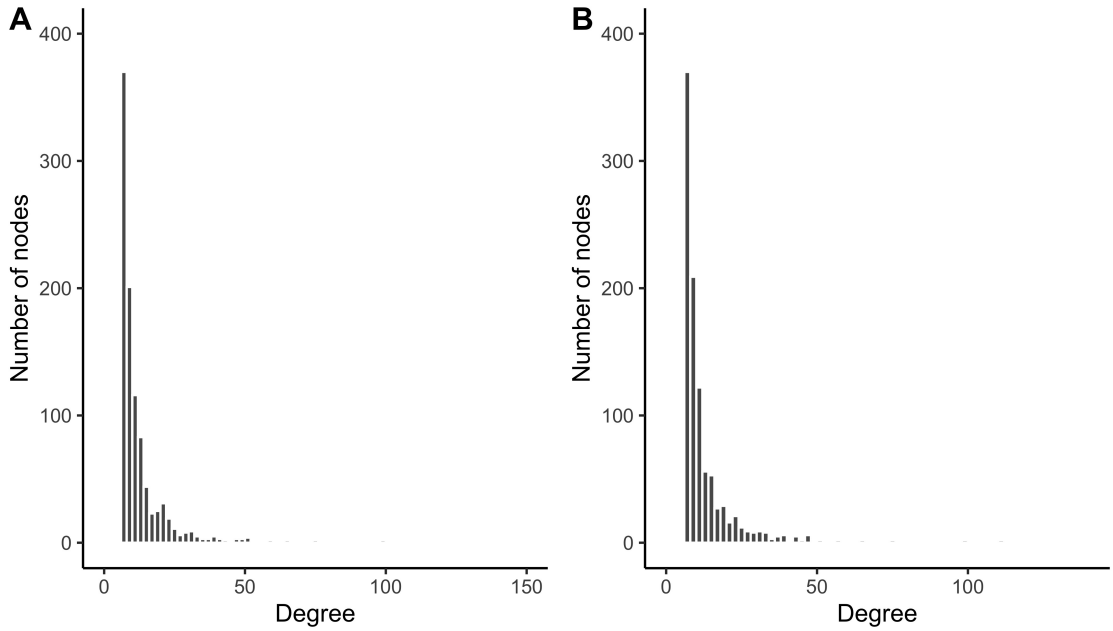

Figure 5.3: Degree distribution of Barabasi-Albert preferential attachment network of one thousand nodes generated with $m = 7$. (A) Barbasi-Albert network 1, (B) Barbasi-Albert network 2.

While the degree distribution of the Erdos-Renyi networks is not all that different from our multilayer network, the edge weights are modified. In the multilayer network, the disease edge weights are based on whether two nodes are in the same household, workplace, or friend

group. In the Erdos-Renyi network, all disease edge weights are 0.71, a weighted average of the household, work, and friend edge weights from the multilayer network. This edge weight is also used in the generated Barbasi-Albert preferential attachment networks.

With these networks in mind, we can now turn to the disease and behavior simulations.

5.3 Model Simulations

To simulate a disease outbreak we initialize the model in the following way. First, we construct the contact network, as detailed above and in Chapter 3. Then, we seed a single node at random and this node will serve as the seed for every outbreak simulation on this network. The seed is symptomatic and infectious (I_2). If behavior is part of the model, the seed is also masking. Three neighbors of the seed are presymptomatic infectious (I_1) and masking, while four different neighbors are exposed (E) and will become symptomatic. These exposed neighbors are not masking at the start of the simulation. This initialization assumes that the seed infects at least seven others, which is somewhat high. However, superspreading events play a role in the transmission dynamics of SARS-CoV-2, so this assumption is not unrealistic [3].

The chance of infection of an individual is based on the disease state of the infectious node, whether or not both nodes are masking, and the strength of the interaction between them:

$$\mathbb{P}(\text{transmission from } i \longrightarrow j) = \beta_i \cdot e_{ij} \cdot m_i \cdot m_j.$$

Here, β_i is the β value associated with the infectious individual's infectious state (I_1, I_2, A_1, or A_2) and e_{ij} is the disease edge weight between individual i and j (based on household, work, or friend interaction). The m_i and m_j terms represent the reduced probability of transmission if the infectious or susceptible individuals mask, respectively. If the infectious node masks, the probability of transmission is reduced by 70%, meaning $m_i = 0.3$. Otherwise, $m_i = 1$. If the susceptible node masks the probability of transmission is reduced by 20%, meaning

$m_j = 0.8$. Otherwise, $m_j = 1$.

All analyses were conducted in R version 4.0.3 and all code will be available on GitHub at `http://github.com/jtaube/taube-bowdoin-honors`.

5.3.1 No Masking Behavior

We first simulate disease propagation across our multilayer network without any masking behavior. These simulations serve as a baseline to which we can compare outcomes of simulations that allow masking and also verify that our network disease model is functioning properly. As expected, the number of newly infected individuals rises and falls as the disease spreads throughout the network (Figure 5.4). Across networks, the total number of infected individuals ranged from 660 to almost 780 (Figure 5.5A). The peak number of daily new infections per outbreak generally fluctuated between 20 and 27 new infections (Figure 5.5B), while the median time at which the outbreak peaked varied between 40 and 55 time steps (Figure 5.5C). All values were relatively consistent across networks.

5.3.2 Mask Mandate

In contrast, we can look at the results of disease outbreaks in populations 85% compliant with mask mandates (850 people in our 1000 node networks). In this case, the seeded infected nodes must also be compliant with the mask mandate. With a mask mandate in place, the total number of infected people rarely exceeds 30, compared to over 650 infections when no one masked (Figure 5.6A). In Figure 5.6B, we see that the daily peak number of new infections is typically only 1 or 2. The timing of these peaks is also much earlier than the outbreaks without masking (Figure 5.6C), with network medians all less than 10 time steps. Ultimately, masking reduces infections in our model, shifting the peak outbreak time earlier.

Outbreaks without masking and with mask mandates lie at the two extremes of reality. We are really interested in the region in between: when people can choose whether to mask, how many will do so? Based on what factors?

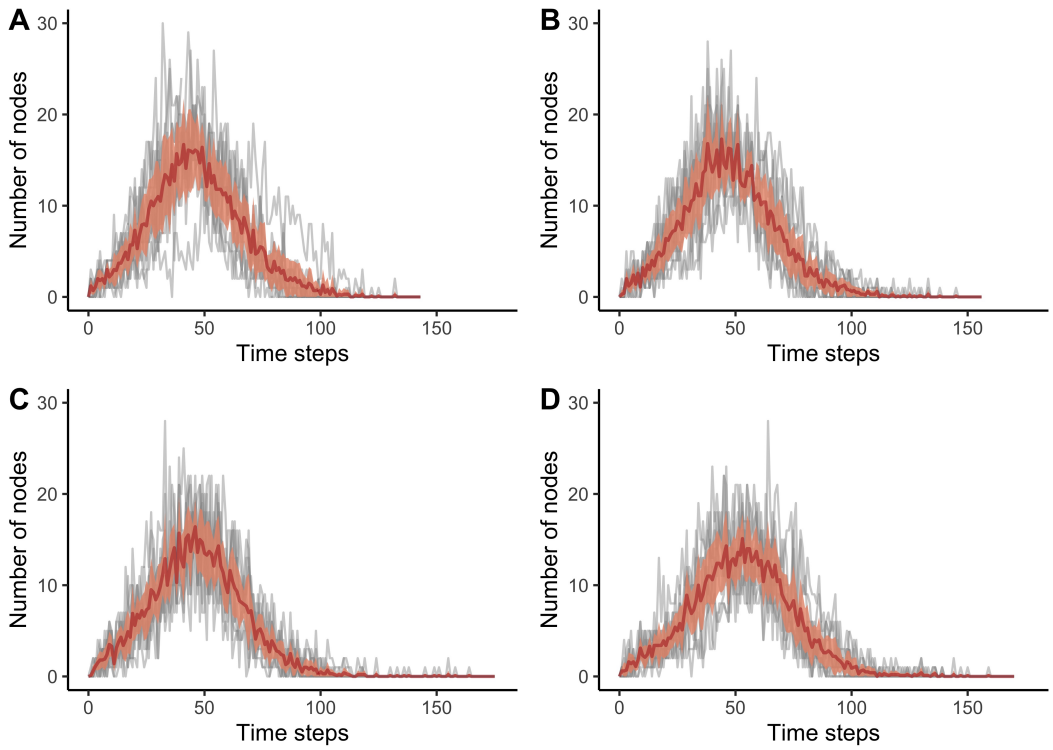

Figure 5.4: No masking behavior. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval.

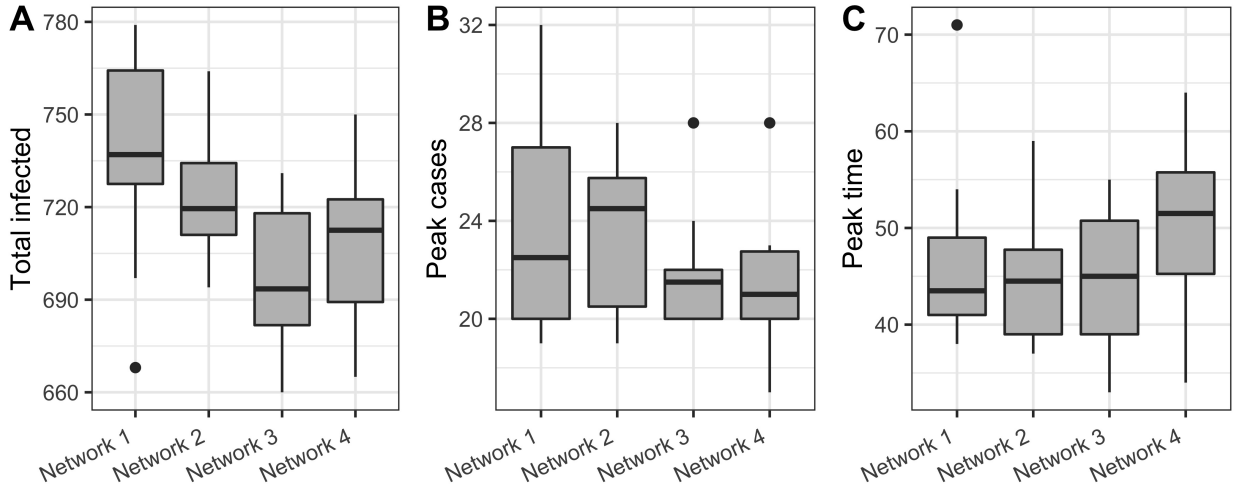

Figure 5.5: No masking behavior. Outbreak statistics for each multilayer network. (A) Total number of nodes infected. (B) Peak number of daily new infections. (C) Time step of peak number of daily new infections.

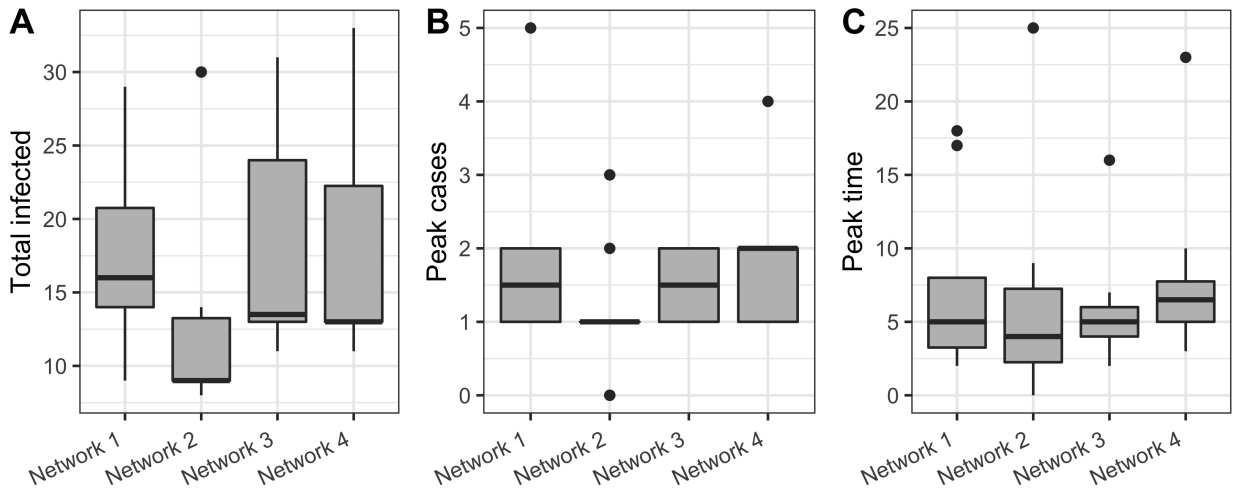

Figure 5.6: Mask mandate with 85% compliance. Outbreak statistics for each multilayer network. (A) Total number of nodes infected. (B) Peak number of daily new infections. (C) Time step of peak number of daily new infections.

5.3.3 Role of Resistance

To address these questions, we begin by looking at our behavior model without resistance ($\tilde{q}_i = 0$ for everyone). In other words, individuals' decisions about masking are solely based on their neighborhood risk and social influence. One concern in a model without masking resistance is that people may continue to mask even when there is no risk of infection. For example, if everyone is masking and disease risk is eliminated, social influence will continue to encourage masking. This behavior would not be realistic. We will see later in section 5.3.5 why this doesn't happen, but this possibility initially motivated our use of resistance in the model.

In Figure 5.7, we see that in the absence of resistance, the number of people masking dramatically dominates the number of people infected. Further, the peak of the average masking curve (in blue) roughly aligns with the peak of the average infection curve (in red). No more than around 200 people mask on a given day (Figure 5.7) and rarely do more than 330 total different people mask (Figure 5.8A). Fewer total maskers leads to more infections than in the mask mandate model (Figure 5.8B), though still nearly 100 fewer infections

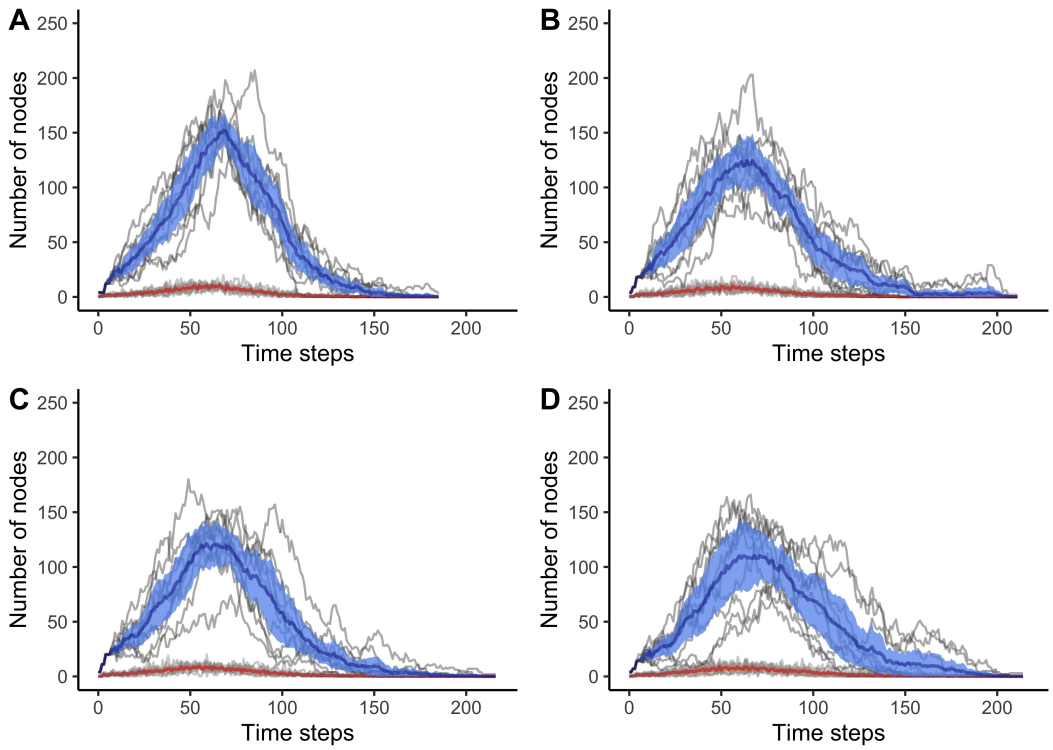

Figure 5.7: No resistance to masking; only perceived risk and social influence contribute to masking decision. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval.

than outbreaks without masking. Peak daily cases and time also fall in between the mask mandate and no masking models, though closer to the no masking simulation runs (Figure 5.8 C, D).

When we include resistance to masking in the model, we must consider its scale – or importance – relative to social influence and perceived risk. Since we have no data quantifying resistance to masking, it's even more important to use the model to explore the potential interplay between resistance and our other factors impacting behavior. The relative scale of resistance is represented by the factor L, where larger values of L indicate higher levels of resistance relative to risk and social influence. If resistance is too high, it can overpower social influence and perceived risk so that no one masks. To explore an appropriate scale of

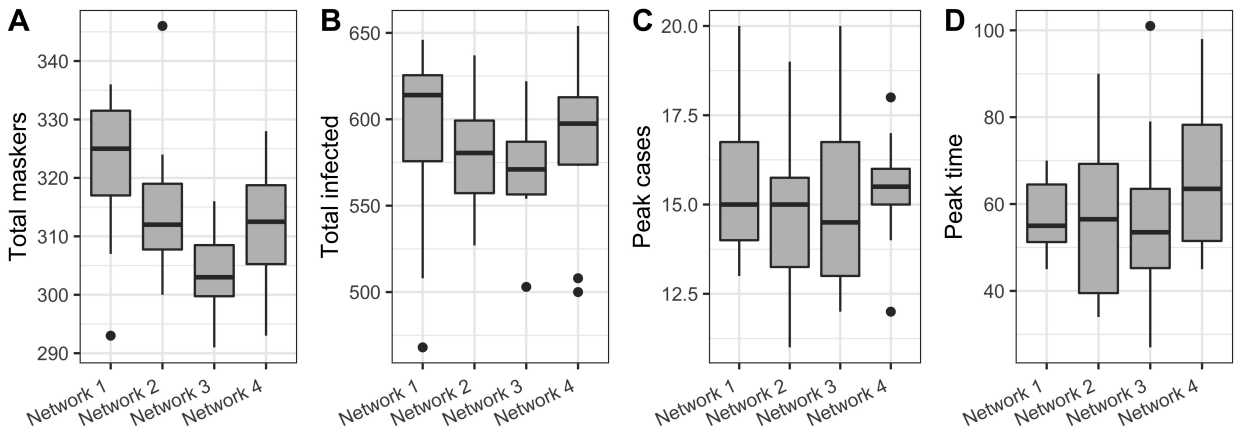

Figure 5.8: No resistance to masking; only perceived risk and social influence contribute to masking decision. Outbreak statistics for each multilayer network. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections.

resistance we run the equilibrating behavior model with L values of 0.01, 0.1, 0.2, 0.4, 0.8, and 1. The results of L values of 0.01, 0.1, and 0.2 are shown in Figures 5.9 and 5.11 and the results of L values of 0.4, 0.8, and 1 are shown in Figures 5.10 and 5.12.

With increasing resistance, the total and daily number of masking individuals decreases substantially. Median total maskers is over 300 for all four networks when maximum resistance is 0.01 but below 100 when maximum resistance is 1 (Figures 5.11 A, 5.12 A). Trends in total number of infected individuals, peak cases, and peak times are less notable across the resistance levels aside from the increased number of infections when fewer individuals mask. We feel that any maximum resistance between 0.1 and 0.4 is reasonable, as these L values don't completely prevent masking but are large enough to compete with influence and risk terms. We will be using 0.2 as the maximum resistance in the remaining experiments.

As a baseline, Figure 5.13 shows again the 10 simulations on each multilayer network with the behavior model using maximum resistance of 0.2. The summary outbreak statistics are shown in Figure 5.14.

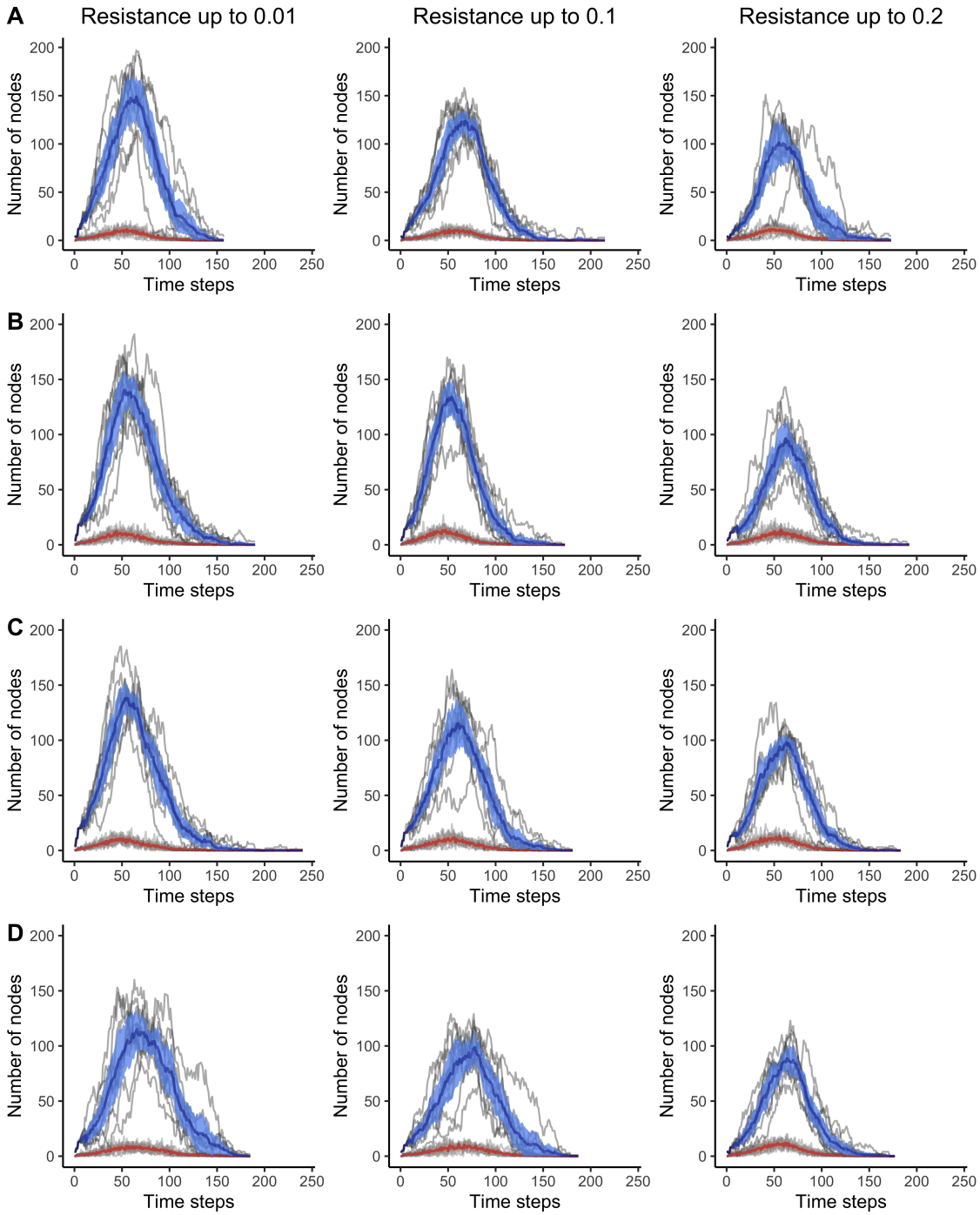

Figure 5.9: Varying maximum resistance to masking. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each column is a maximum resistance level, 0.01, 0.1, and 0.2, from left to right. Each row is a network; network 1, 2, 3, and 4 from top to bottom. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval.

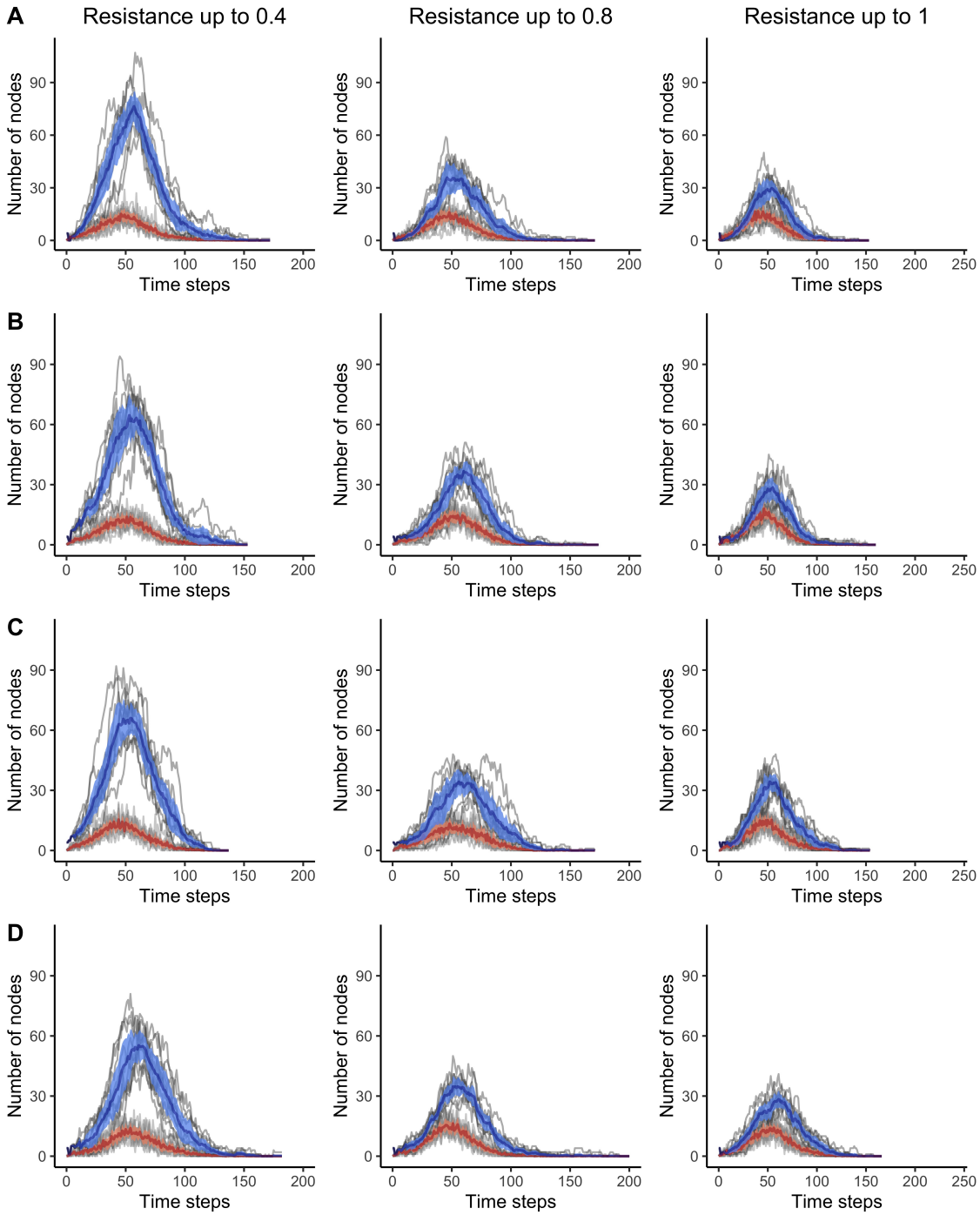

Figure 5.10: Varying maximum resistance to masking. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each column is a maximum resistance level, 0.4, 0.8, and 1, from left to right. Each row is a network; network 1, 2, 3, and 4 from top to bottom. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval.

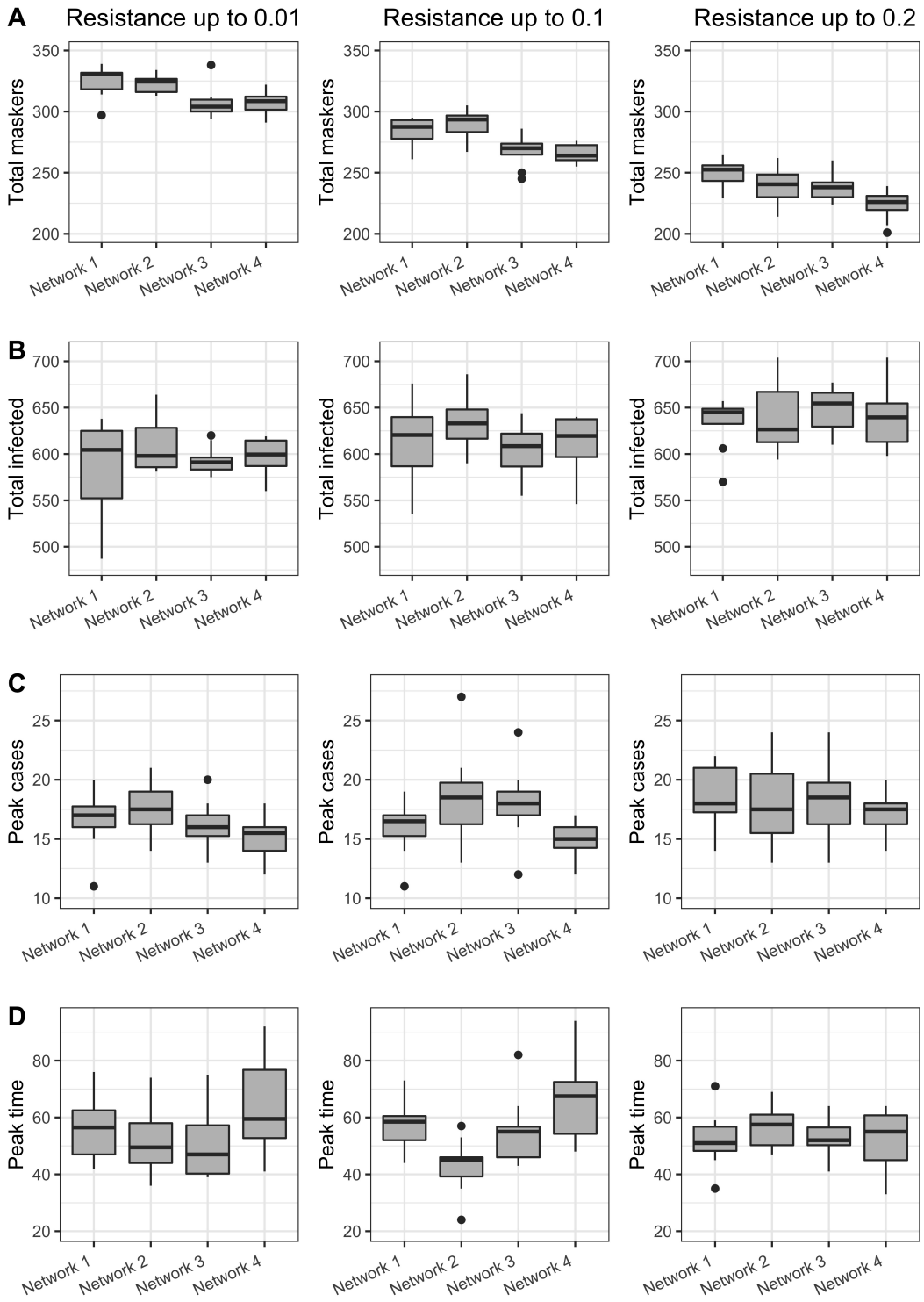

Figure 5.11: Varying maximum resistance to masking. Outbreak statistics for each multilayer network. Each column is a maximum resistance level, 0.01, 0.1, and 0.2, from left to right. Each row is a statistic. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections.

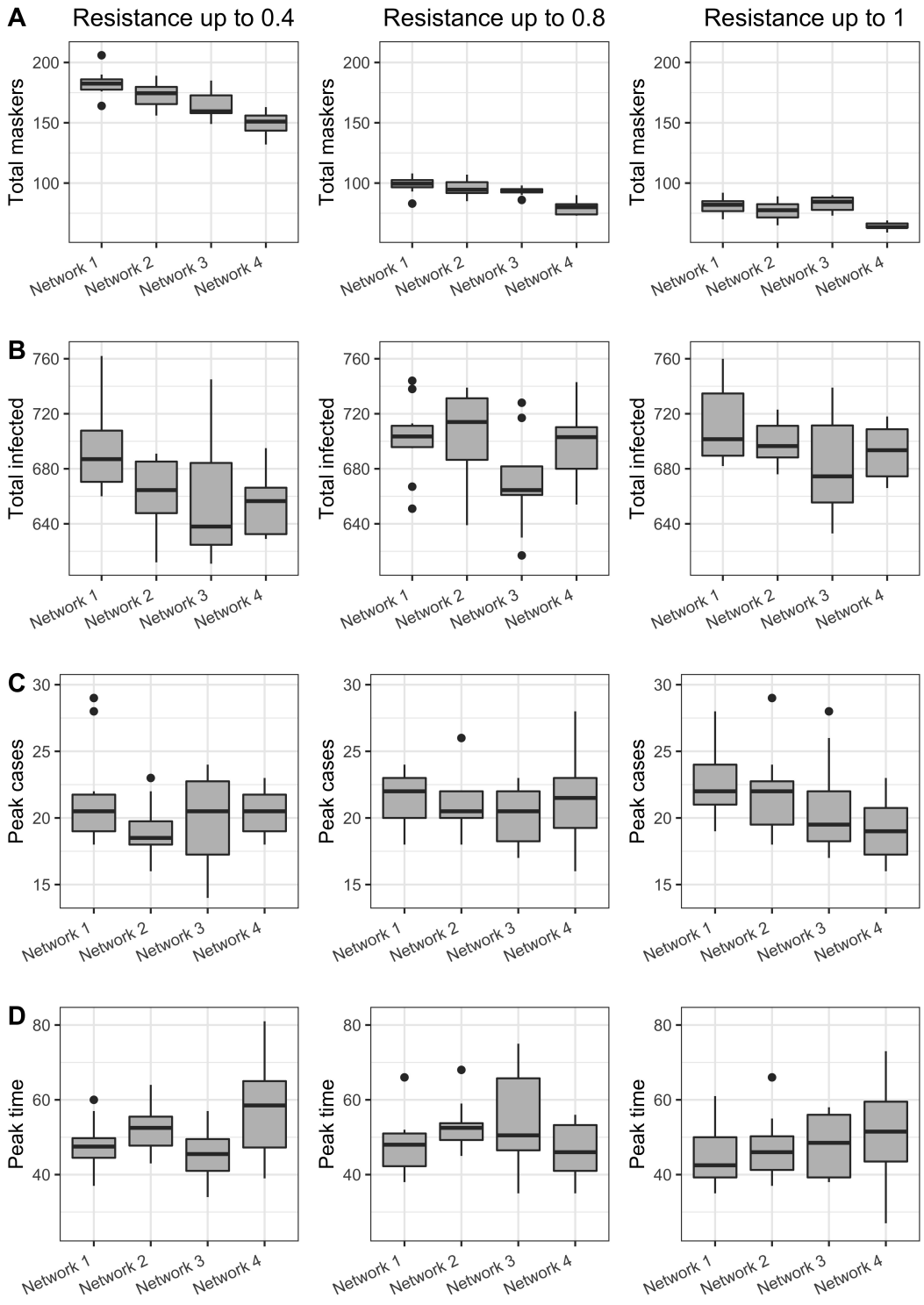

Figure 5.12: Varying maximum resistance to masking. Outbreak statistics for each multilayer network. Each column is a maximum resistance level, 0.4, 0.8, and 1, from left to right. Each row is a statistic. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections.

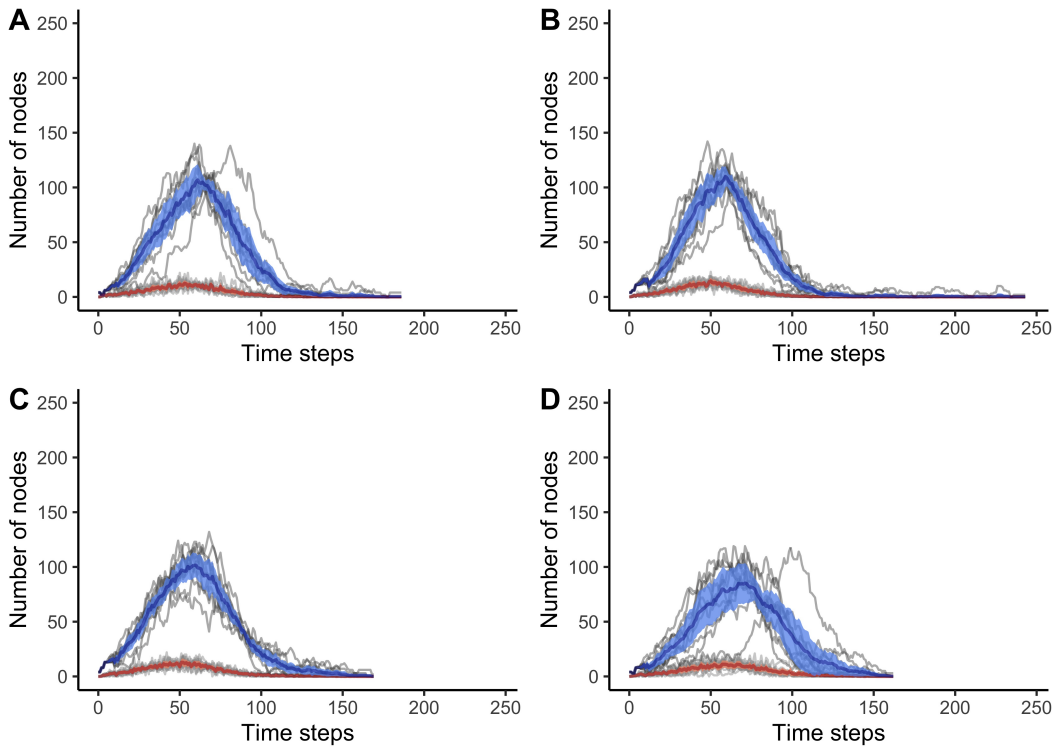

Figure 5.13: Maximum resistance of 0.2. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval.

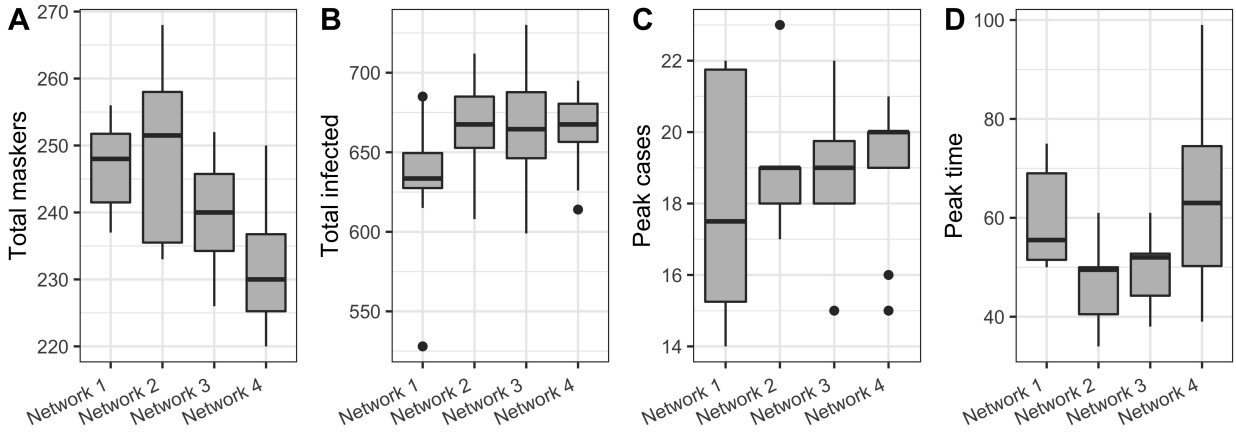

Figure 5.14: Maximum resistance of 0.2. Outbreak statistics for each multilayer network. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections.

5.3.4 Role of Perceived Risk

Fear is thought to be a large driver of precautionary behavior, especially in the context of COVID-19 [36, 48]. We further investigate the role of perceived risk in our model by eliminating it all together. Will anyone mask if no one perceives any risk of disease and everyone follows what their neighbors are doing? The way we've initialized our model with only a few maskers in a network of one thousand, early on there is overwhelming social influence not to mask. Thus, in the absence of perceived risk no one masks and we observe an epidemic curve as we saw when we modeled disease without behavior (Figures 5.15 and 5.16). The four maskers in Figure 5.16A are those that were initialized as masking in the model. Evidently, risk is the main driver of masking behavior in our model.

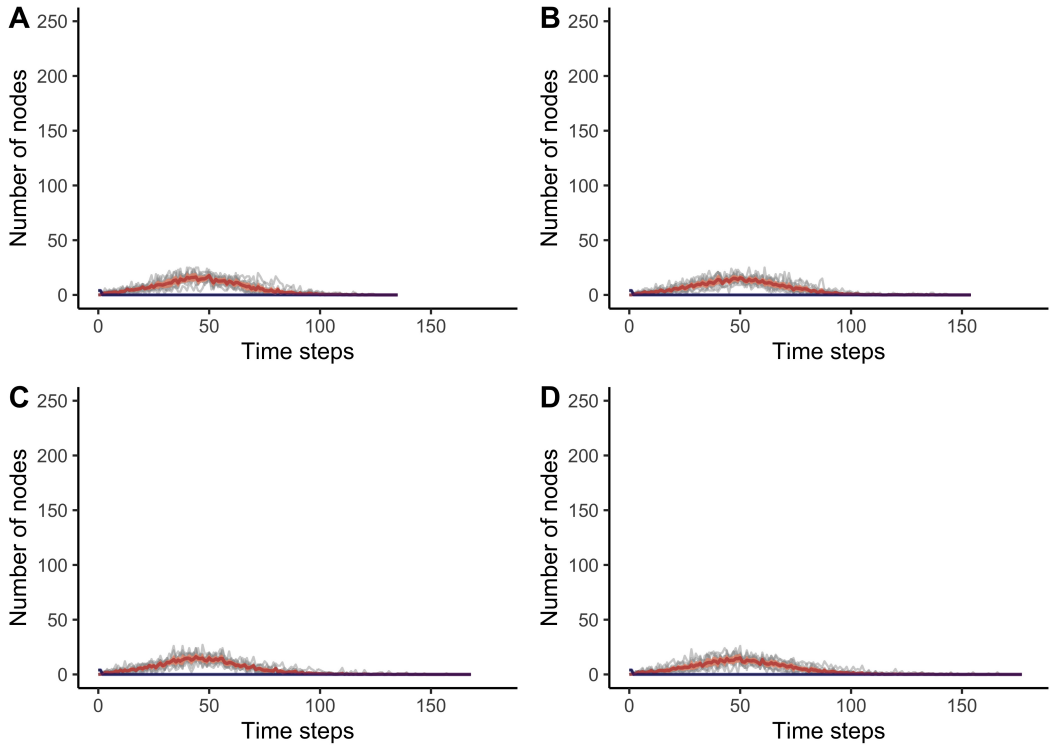

Figure 5.15: No perceived risk of infection; only resistance (up to 0.2) and social influence contribute to masking decision. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes.

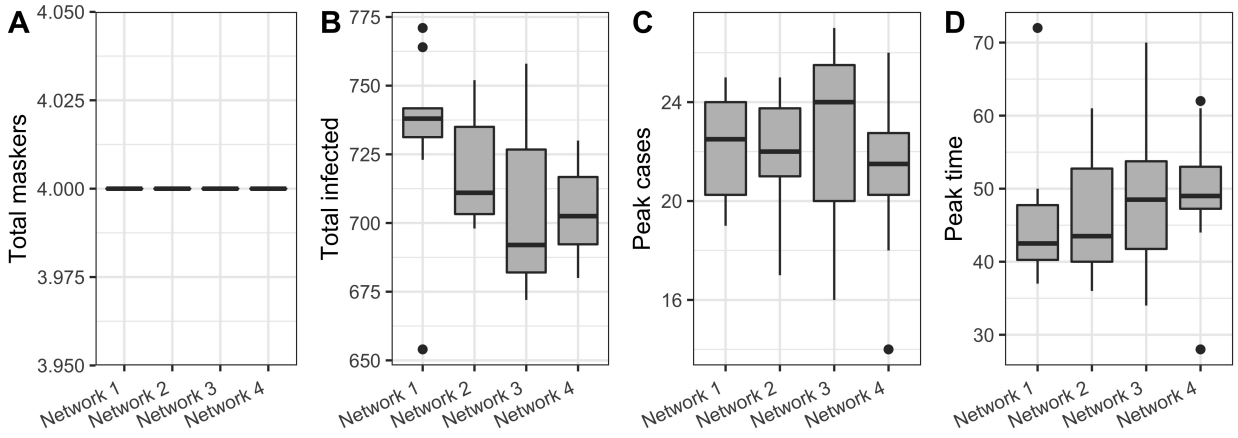

Figure 5.16: No perceived risk of infection; only resistance (up to 0.2) and social influence contribute to masking decision. Outbreak statistics for each multilayer network. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections.

In our calculation of perceived risk, we don't account for how masked neighbors may reduce a node's risk of infection. If we include neighbor masking behavior in the risk calculation, we may see some kind of free rider effect where a node with many masking neighbors perceives that they are at lower risk of infection and, thus, they are less inclined to mask. We test this idea by scaling perceived risk by the proportion of masking neighbors. Let the set U_i contain all of i's neighbors and the set M contain all masking individuals. If 0.7 is the reduction in the probability of transmission from the infectious individual masking, then the new term in our risk product is

$$1 - 0.7\frac{|M \cap U_i|}{|U_i|} \tag{5.1}$$

where $|.|$ denotes cardinality. We include the 0.7 scalar in this expression to highlight that masking does not completely eliminate the risk of transmission. When we incorporate expression 5.1 into our perceived risk formula we have

$$\rho_i = \sqrt{c_i \cdot \frac{|(A_1 \cup A_2 \cup I_1 \cup I_2) \cap U_i|}{|U_i|}} \cdot \Big(1 - 0.7\frac{|M \cap U_i|}{|U_i|}\Big). \tag{5.2}$$

In equation 5.2, we can see that fewer masking neighbors lead to larger values of expression 5.1 and, consequently, larger perceived risks of infection, whereas more masking neighbors reduce perceived risk.

When we implement this model, we see levels of masking similar to that seen in the original behavior model (Figures 5.17 and 5.18).

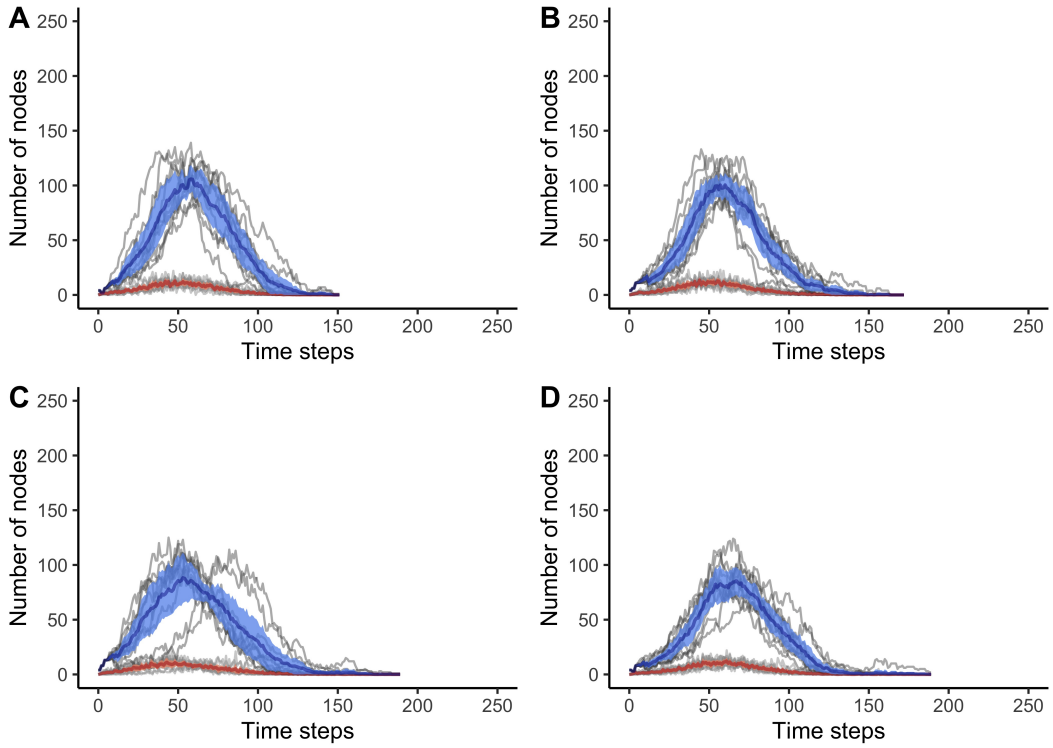

Figure 5.17: Perceived risk incorporating proportion of masking neighbors. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval.

The median total number of maskers ranged from 230 to 255 in the original behavior model, and with this new risk structure the median total number of maskers vary from 220 to 255, with the lowest occurring in network 4. Likewise, the median total number of infections appear to be about the same or a bit lower with this new risk structure than in the original behavior model. Scaling perceived risk by neighbors' masking decisions did not produce the free rider effect we had expected, nor did this risk structure change the

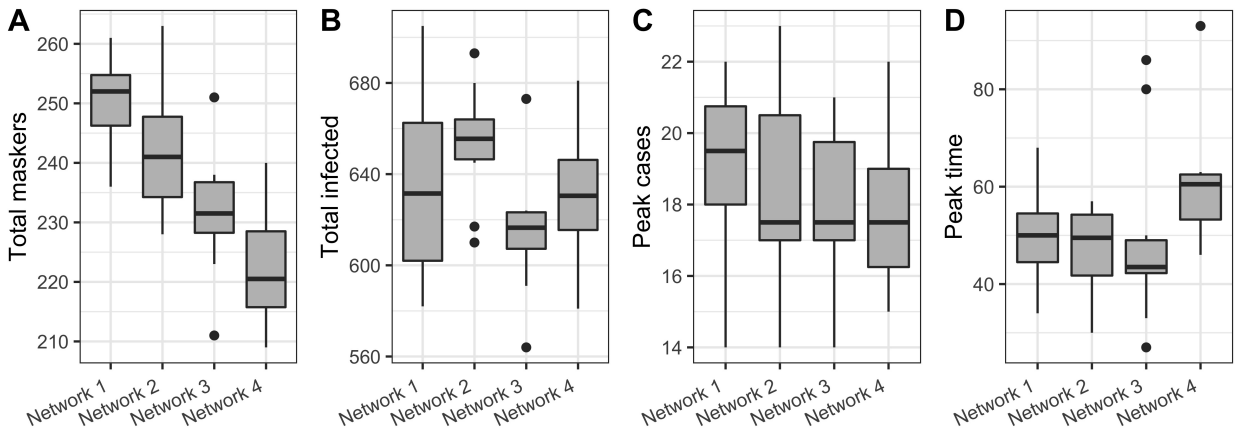

Figure 5.18: Perceived risk incorporating proportion of masking neighbors. Outbreak statistics for each multilayer network. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections.

overall masking and disease dynamics. This result indicates the robustness of our original risk formulation, though additional testing for the free rider effect when masking is initially more widespread is necessary.

5.3.5 Role of Social Influence

We also run the equilibrating model without social influence to see how individuals behave when only considering disease risk and masking resistance. When social influence isn't present, the number of total maskers in each outbreak ranges from under 200 to over 800, compared to between 220 and 270 total maskers in the original behavior model (Figure 5.20D vs. 5.14D). The large number of maskers drives down the total number of cases, the peak new cases per day, and often the peak day of the outbreak (Figure 5.20). Number of maskers per day increases initially as people have a chance to respond to the seed infections in their neighborhood, and then, on average, declines toward zero as the outbreak wanes (Figure 5.19).

This average seems to be obscuring a lot of the variation across simulations. In Figure 5.21 we look more closely at how the number of people masking each day fluctuates over the

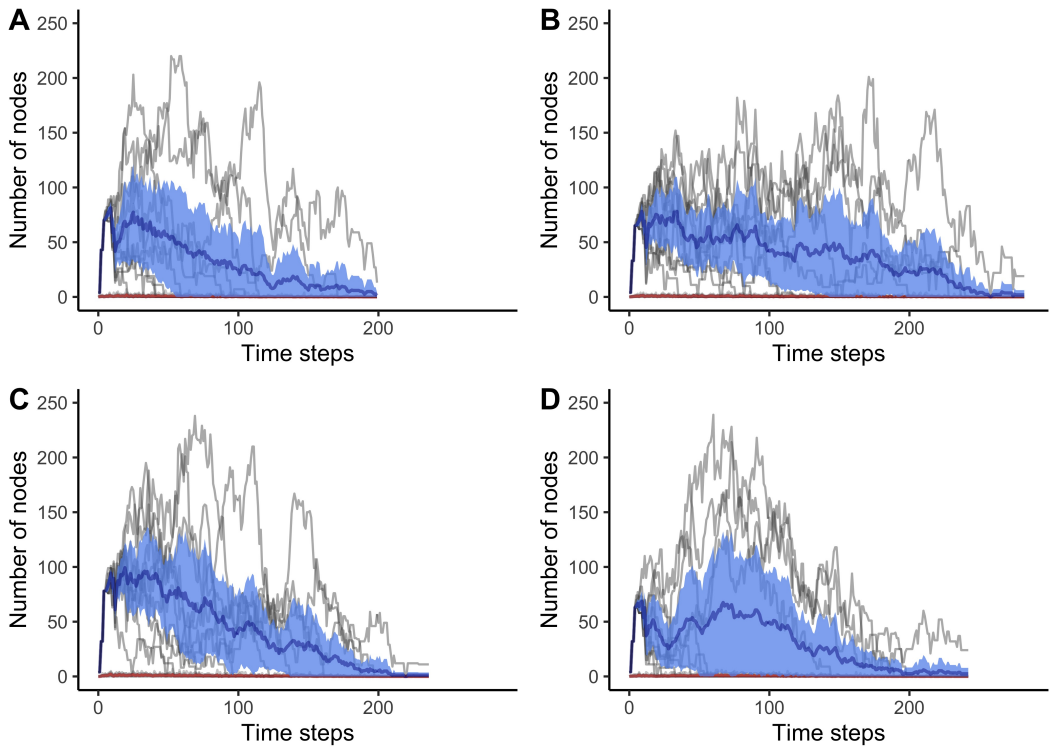

Figure 5.19: No social influence; only resistance (up to 0.2) and perceived risk contribute to masking decision. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval.

course of the outbreak across ten simulations on multilayer network 1. We observe multiple waves of increasing maskers per day in the light purple, blue, and green curves. These trends contrast with what we have seen in our model that includes social influence, where the number of daily maskers curves typically have a single peak. The rise and fall of masking observed in the absence of social influence aligns more with what we had expected to see, in that as infections increase, masking increases, then infections fall, and masking falls, and so on. These results demonstrate how much social influence is *preventing* people from masking and stabilizing masking behavior in our model. The lack of stability without social influence may be unsurprising given that when we eliminate social influence, there is no longer an equilibration process. In the original behavior model, the only factor changing during the

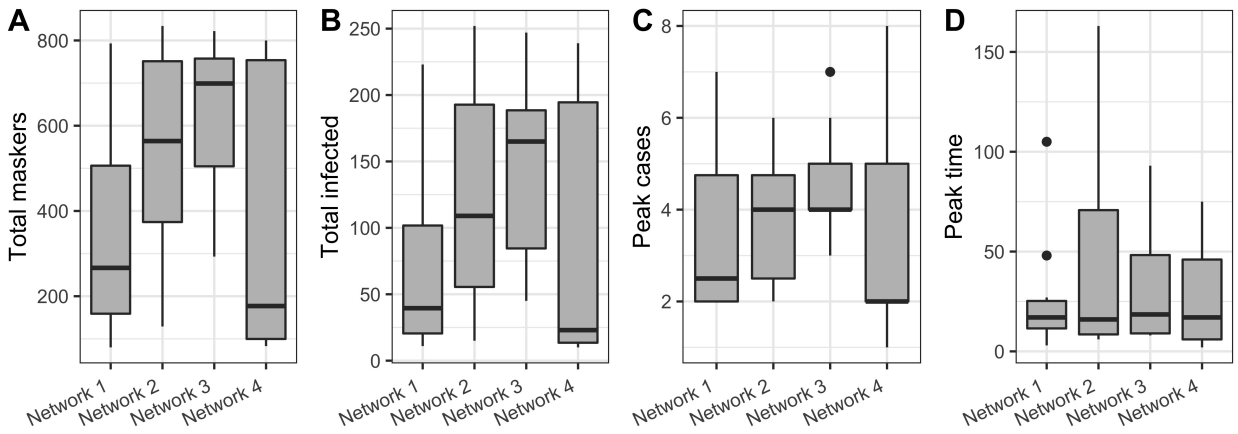

Figure 5.20: No social influence; only resistance (up to 0.2) and perceived risk contribute to masking decision. Outbreak statistics for each multilayer network. (A) Peak number of daily new infections. (B) Time step of peak number of daily new infections. (C) Total number of nodes infected. (D) Total number of nodes who masked at some point during the outbreak.

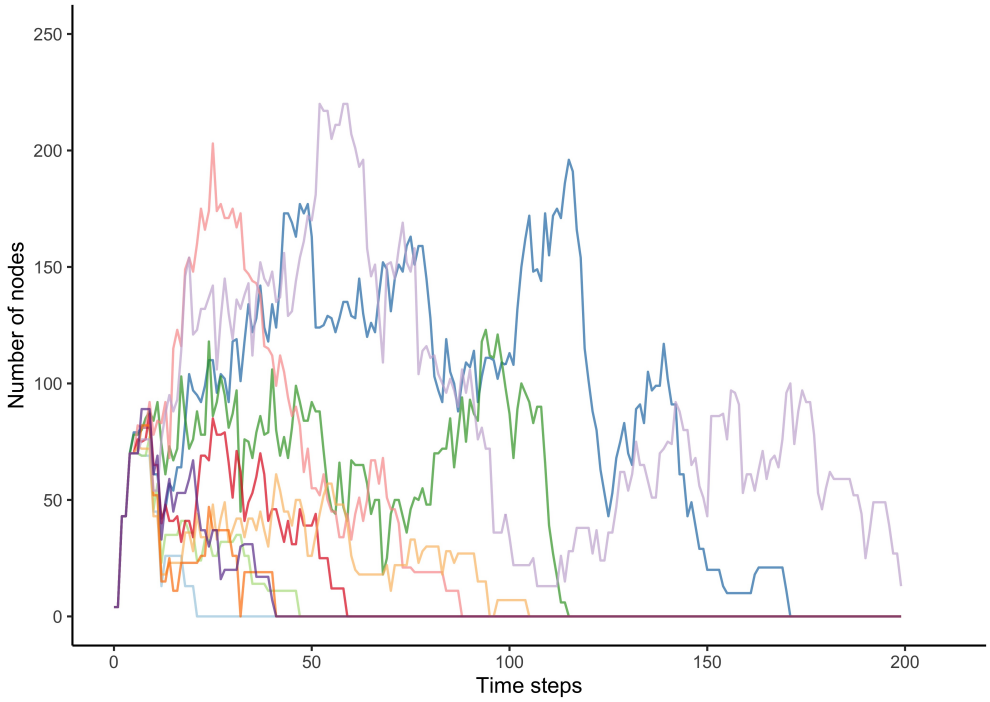

Figure 5.21: No social influence; only resistance (up to 0.2) and perceived risk contribute to masking decision. Number of masking nodes per day shown for 10 different simulations on multilayer network 1 (1000 nodes). Each simulation is a different color. Note that the light purple, blue, and green curves appear to have multiple waves of increased number of maskers per day.

equilibration process was the behavior of one's neighbors – neither masking resistance nor perceived risk changed. Thus, removing social influence means no equilibration occurs in the decision-making process. We will further explore the role of equilibration in the next section, but first, we look at the relative importance of risk and social influence.

Risk will always push people to mask, while resistance will always prevent people from masking. Social influence can go both ways but in our model appears to mostly act against masking behavior. Factors against masking (resistance and perhaps influence) are often outweighing factors encouraging masking (risk and perhaps influence) in our original model. We must consider that risk and social influence may not equally contribute to the decision whether to mask.

To explore this possibility, we experiment with maximum incoming influence values. In our original behavior model, the sum of incoming influence weights, ϕ, was uniformly drawn from [0,1]. Here we test ϕ values drawn uniformly from [0, 0.1], [0, 0.25], [0, 0.5], and [0, 0.75], in addition to [0, 1]. With incoming influences drawn from [0, 0.1], [0, 0.25], or [0, 0.5], the masking dynamics appear very similar to the model without social influence: on average, the number of masking individuals rises quickly at the beginning of the outbreak and then decreases more slowly back to zero (Figures 5.22 and 5.23). With ϕ values drawn from [0, 0.75] we begin to see the tension between social influence and perceived risk. The average peak of the number of maskers per day curve is highest, likely because social influence is strong enough to synchronize behavior changes and reduce variability but cannot completely dominate perceived risk which increases the number of maskers. It also appears that, on average, the number of daily maskers rises for a longer period of time and then decreases more quickly when ϕ can range up to 0.75 versus lower maximum ϕ values. In the future, we would like to explore the sensitivity of the model to the range of ϕ and look to quantify this tension between perceived risk and social influence.

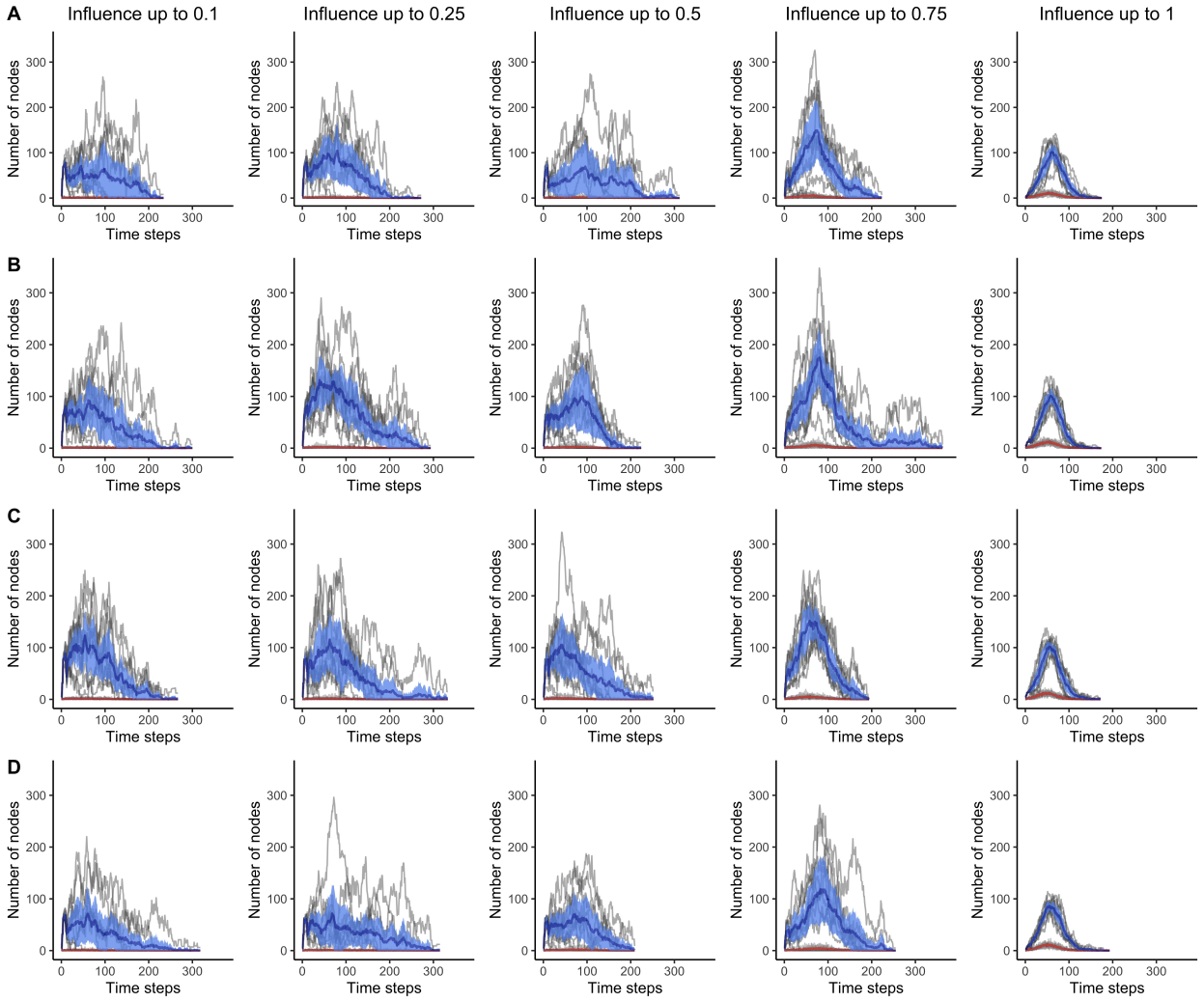

Figure 5.22: Varying maximum incoming social influence. Each column is a maximum influence level, 0.1, 0.25, 0.5, 0.75, and 1, from left to right. Each row is a network; network 1, 2, 3, and 4 from top to bottom. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval.

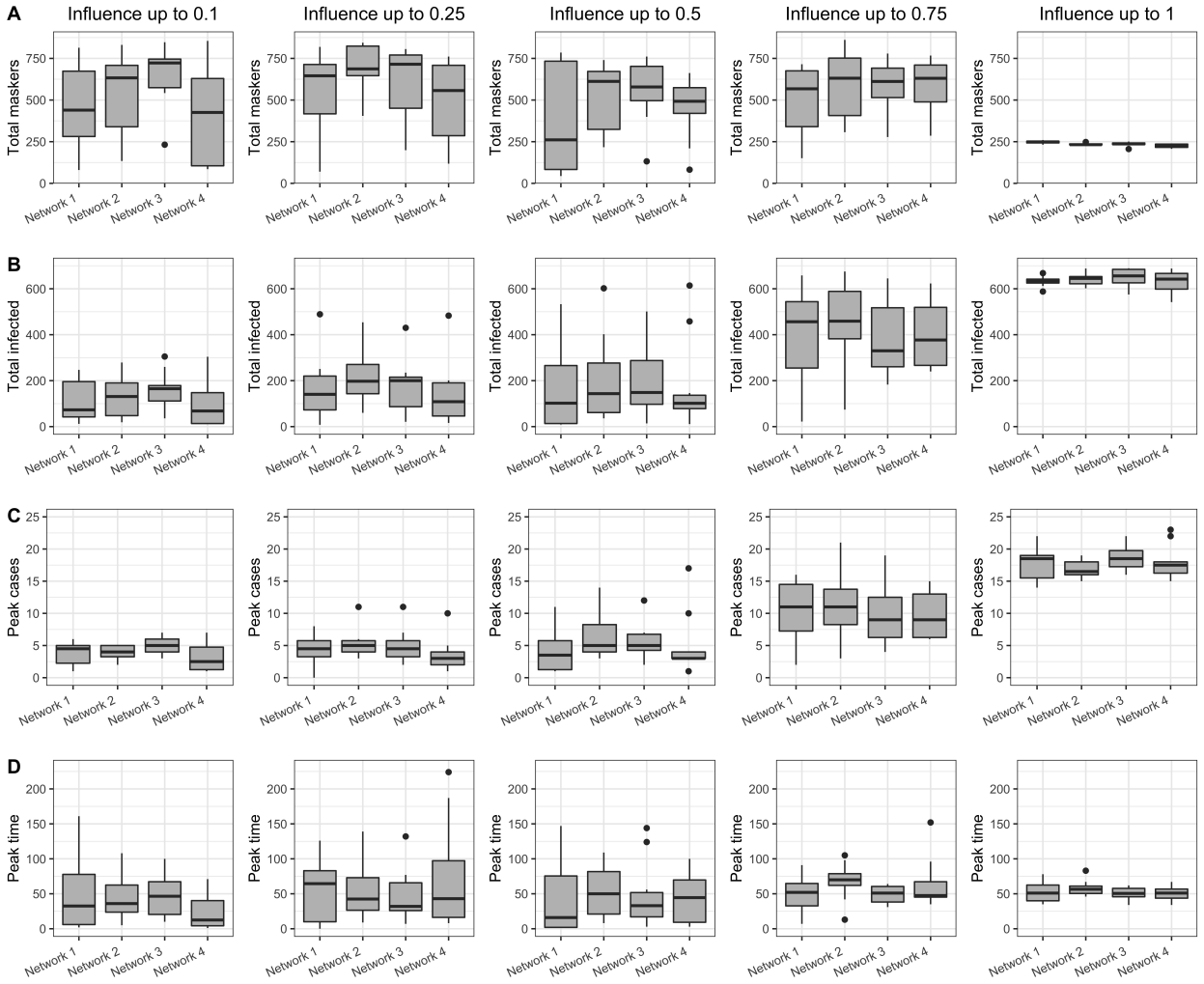

Figure 5.23: Varying maximum incoming social influence. Outbreak statistics for each multilayer network. Each column is a maximum influence level, 0.1, 0.25, 0.5, 0.75, and 1, from left to right. Each row is a statistic. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections.

5.3.6 Role of Equilibration

Equilibration plays a key role in the behavior dynamics that we observe, as previewed when we removed social influence. Without equilibration, the range of total maskers spans from under 400 to nearly 1000 (the whole network) (Figure 5.25). Due to the greater levels of masking in the non-equilibrating model, the total number of nodes infected and the peak daily cases tend to be lower, and the peak time earlier (Figure 5.25). We can also see in Figure 5.24 how the number of maskers on a given day rises and falls quickly. This phenomenon is likely a result of the switching back and forth that can occur due to social influence on a node being positive in one step and negative the next. The disease is spreading at the same rate as masking decisions are being made, so that may also add variability to masking decisions over time. On the other hand, if we let the behavioral system reach equilibrium without changing the perceived risk during the decision process, we eliminate the switching. Individuals are more forward-looking as they speculate on what others may do and determine the best decision for themselves.

Equilibration often only requires three (or sometimes just two) rounds. What happens if equilibration is always limited to two rounds? In Figures 5.26 and 5.27 we explore this question. Immediately, we can see how with only two rounds of equilibration (only one more than in the no equilibration scenario) we have eliminated the dramatic increases and decreases in number of daily maskers seen without equilbration. In fact, the dynamics and summary statistics aren't all that different from our full equilibration model. These results suggest that two rounds of equilibration are all that is necessary to limit large oscillations in masking behavior.

5.3.7 Role of Network Structure

Next, we will explore the role that our multilayer network structure may have had on the disease-behavior dynamics of our system. Let us first look at the two Erdos-Renyi random

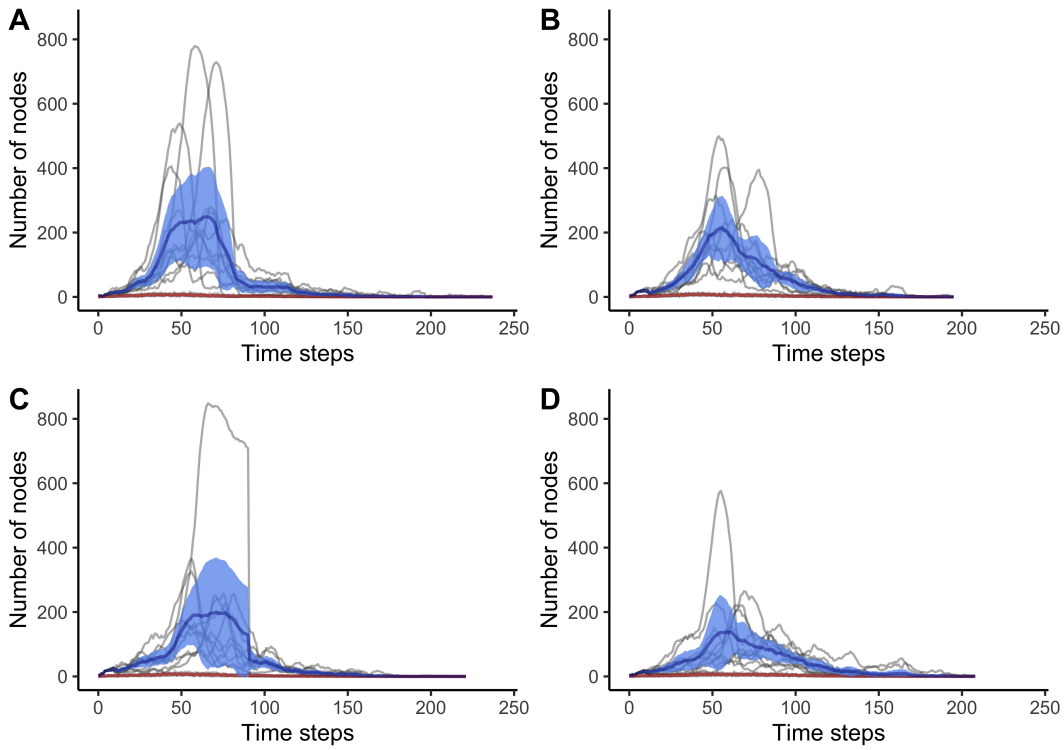

Figure 5.24: No equilibration in the decision-making process. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval.

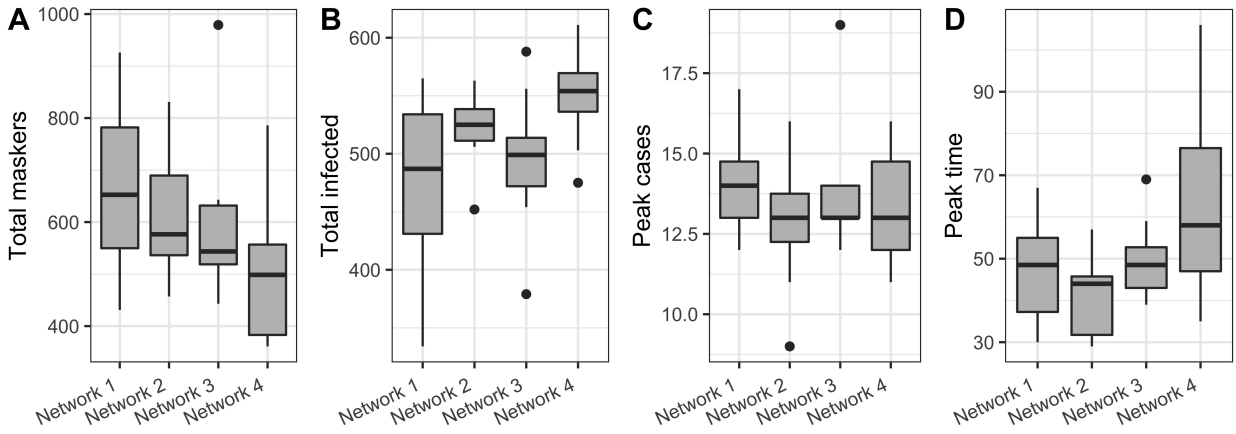

Figure 5.25: No equilibration in the decision-making process. Outbreak statistics for each multilayer network. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections.

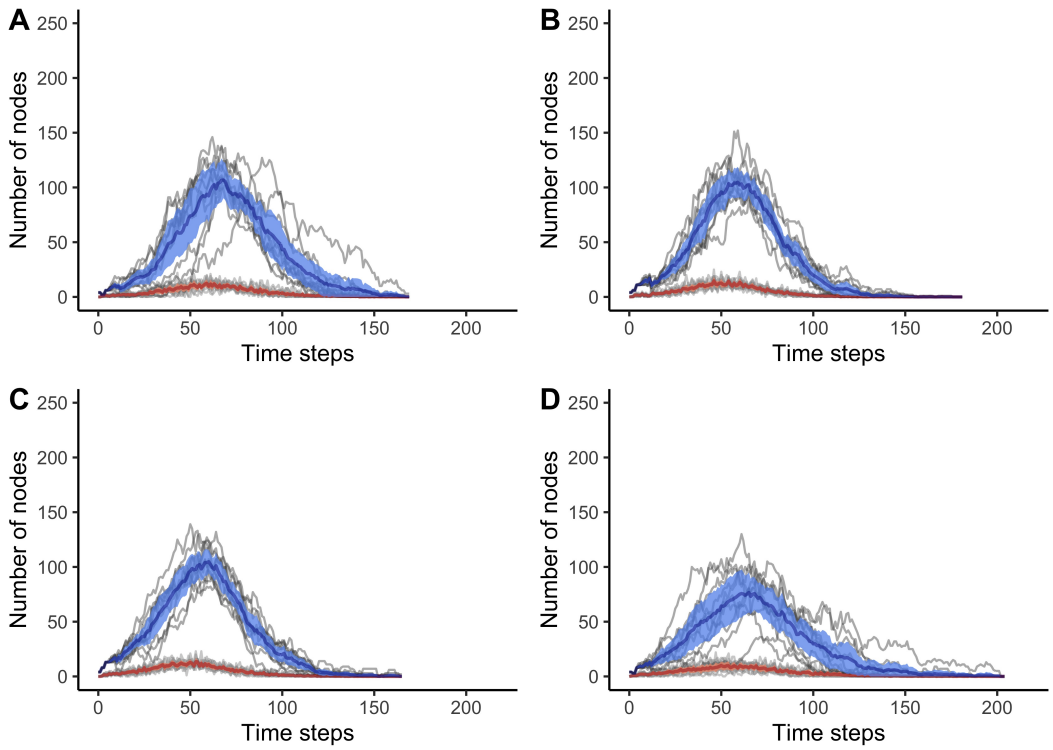

Figure 5.26: Two rounds of equilibration in the decision-making process. Number of newly infected nodes per time step for four one thousand node multilayer networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval.

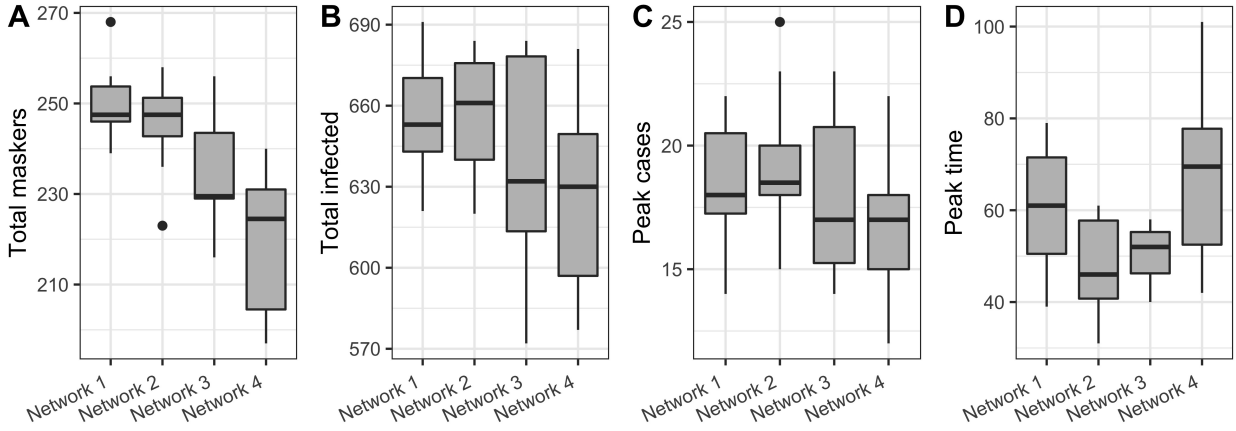

Figure 5.27: Two rounds of equilibration in the decision-making process. Outbreak statistics for each multilayer network. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections.

networks on which we ran the equilibrating behavior model with maximum resistance $L = 0.2$ (Figure 5.28). Recall that these networks are constructed with a constant probability of an edge between any two nodes, in our case $p = 0.014$. Comparing the Erdos-Renyi networks with our multilayer networks, we see slightly shorter peak times, slightly higher totals for infected individuals and peak cases, and roughly the same total number of maskers (Figure 5.29). The shorter peak times and higher number of people infected may be explained by Erdos-Renyi's adherence to the small-world property. Lower clustering of the Erdos-Renyi network compared to the multilayer network would allow the disease to spread more quickly and reach more individuals [35].

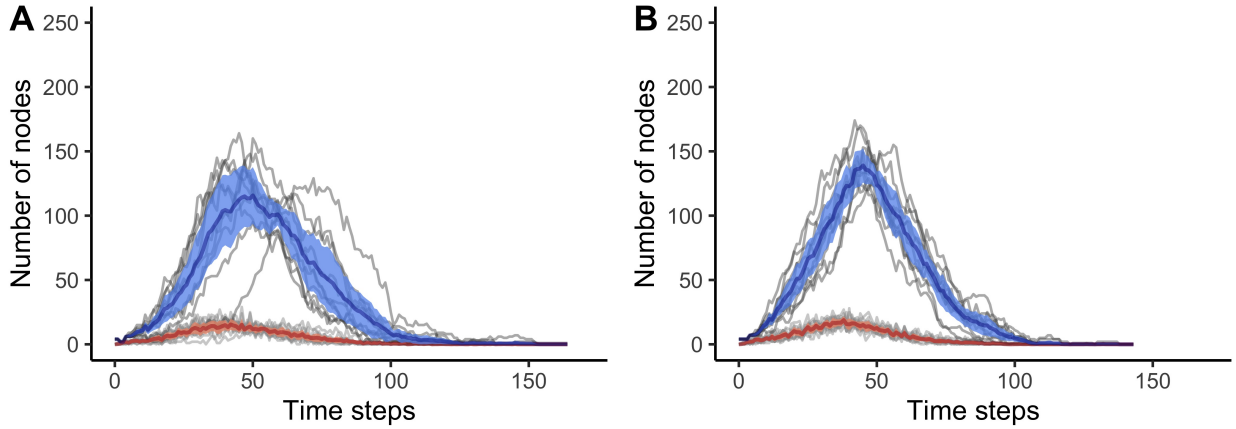

Figure 5.28: Number of newly infected nodes per time step for two one thousand node Erdos-Renyi random networks. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval.

The Barabasi-Albert preferential attachment networks had very different degree distributions than our multilayer network. The power law degree distribution means that a few nodes have very high degrees (sometimes called hubs), while most nodes have lower degrees. In this experiment, we use two different Barabasi-Albert networks. Initially, we seed each at random and then run the outbreak. We then repeat the experiment by instead seeding the highest degree node in the network (> 100 neighbors in these cases). Finally, we seed one of the nodes of lower degree (specifically, degree 7). Interestingly, the results aren't that differ-

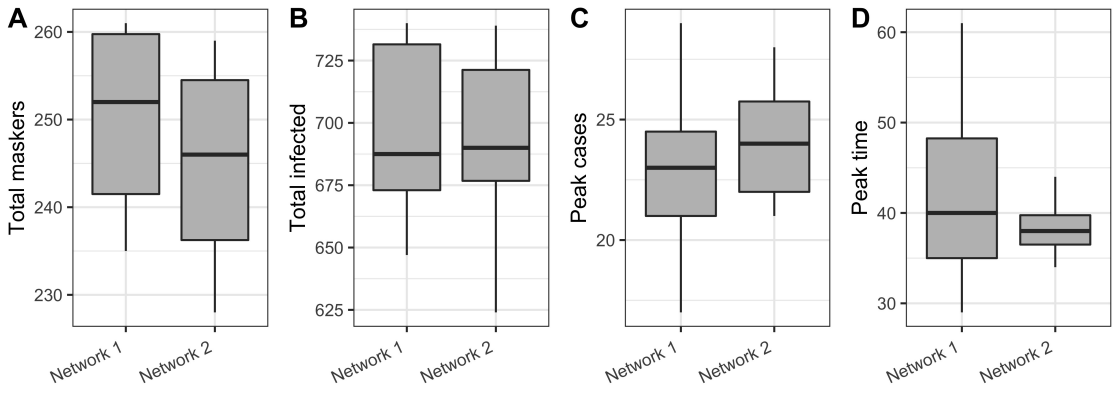

Figure 5.29: Outbreak statistics for two one thousand node Erdos-Renyi random networks. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections.

ent regardless of which way we seed the outbreak (Figure 5.30). Across all four simulations, we see a peak of roughly 300 daily maskers that is a few time steps delayed from the daily new infections peak. Based on Figure 5.31A, it appears that total number of maskers is tied more to the network structure than how the outbreak is seeded. The total number of infected individuals and the peak number of daily cases do not vary noticeably across the different types of seeds (Figure 5.31). The place where seed choice matters most is in the peak outbreak time, where the outbreaks seeded with lower degree nodes took noticeably longer to reach their peak compared to the outbreaks seeded randomly or with highest degree nodes (Figure 5.31D). The hub structure of Barabasi-Albert networks facilitates faster spreading of the disease than seen in our multilayer networks (Figure 5.31D). However, it still takes some time for the disease to spread from a low degree seed to higher degree nodes that can propagate the disease more rapidly, hence the delayed peak time when seeding the outbreak in lower degree nodes. Disease outbreaks on the Barabasi-Albert networks had higher peak cases, total infections, and total number of maskers than on the multilayer networks (Figure 5.31).

In sum, these different networks demonstrate how important contact structure is to behavior and disease dynamics.

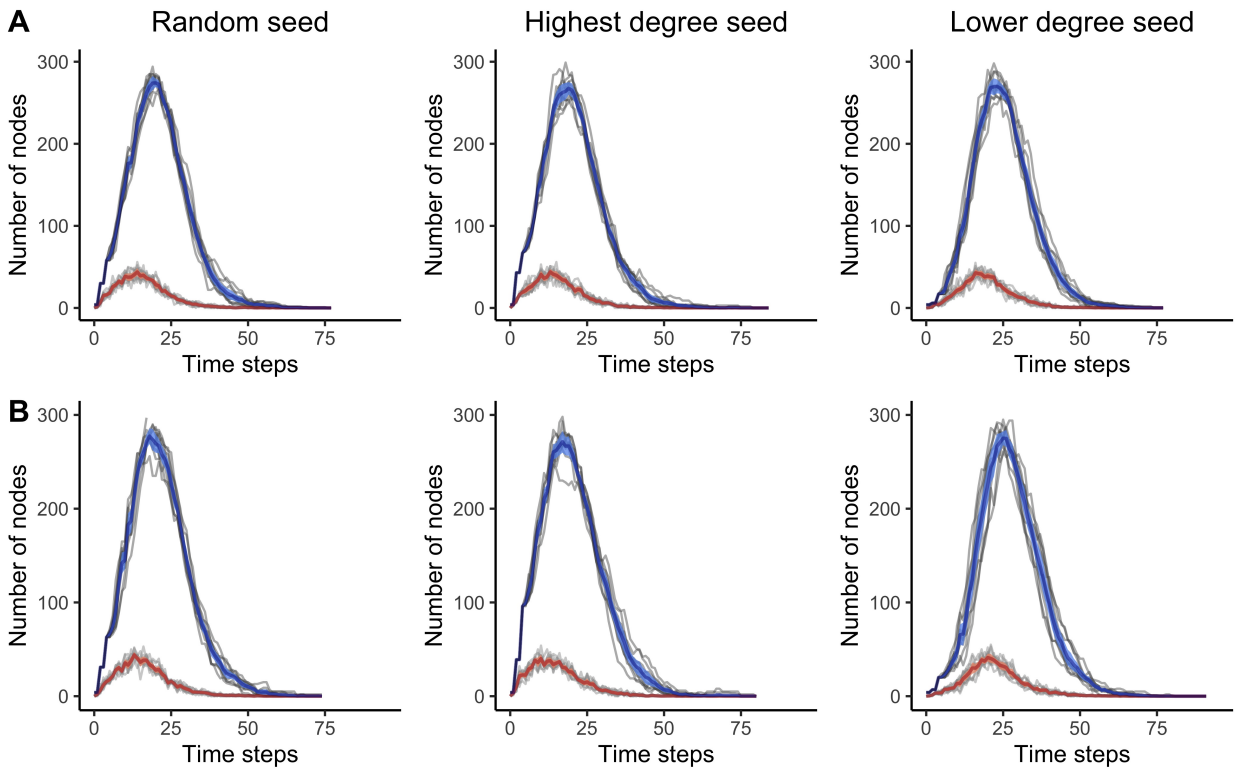

Figure 5.30: Number of newly infected nodes per time step for two one thousand node Barabasi-Albert preferential attachment networks. Columns are seeded differently; randomly, highest degree node, lower degree node, from left to right. Each row is a different network. Each of 10 simulations is shown in grey. Dark red line shows average daily number of newly infected nodes, and pink region shows the 95% confidence interval. Dark blue line shows average daily number of masking nodes, and light blue region shows the 95% confidence interval.

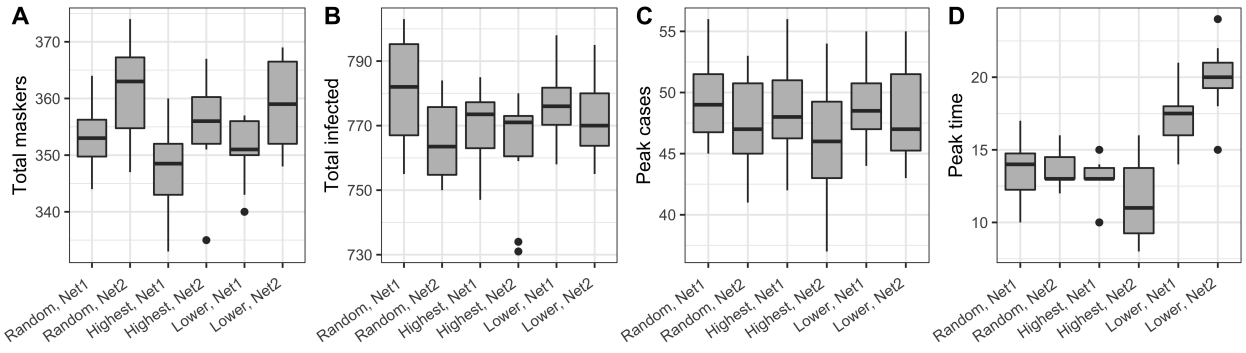

Figure 5.31: Outbreak statistics for each Barabasi-Albert preferential attachment network. Networks are either seeded randomly, with the highest degree node, or with a lower degree node. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections.

Chapter 6

Conclusions

6.1 Key Findings

We built a coupled disease-behavior dynamical model for SARS-CoV-2 using contact networks and drawing from influence games. Individuals are connected within households, workplaces, and friend groups with different interaction strengths, representing the amount of physical contact or duration of time spent together. We based decisions about masking on three factors: individual resistance to masking; perceived risk of infection, based on the proportion of infectious neighbors and personal cost of infection; and social influence from immediate neighbors.

We found that social influence largely prevented masking behavior, whereas risk motivated masking behavior. We used the model to estimate a scale for resistance to masking at which resistance balanced risk in the absence of social influence but played a smaller role when both risk and influence were part of the decision calculation. Equilibration prevented extreme swings in behavior, impacting both masking dynamics and total number of maskers. Moreover, only two rounds of equilibration were necessary to reduce these oscillations. Network structure was also key to masking and disease dynamics: Barabasi-Albert preferential attachment networks had more total maskers and total infected individuals, in addition to

higher peak numbers of daily new infections but substantially lower time until outbreak peak. We can explain these differences by the core-periphery structure of Barabasi-Albert networks. When giving individuals the choice to mask, we rarely saw more than one-third of the network masking at one time, in stark contrast to the mask mandate with 85% compliance that we used as a comparison to outbreaks without masking. These results are summarized in Figure 6.1.

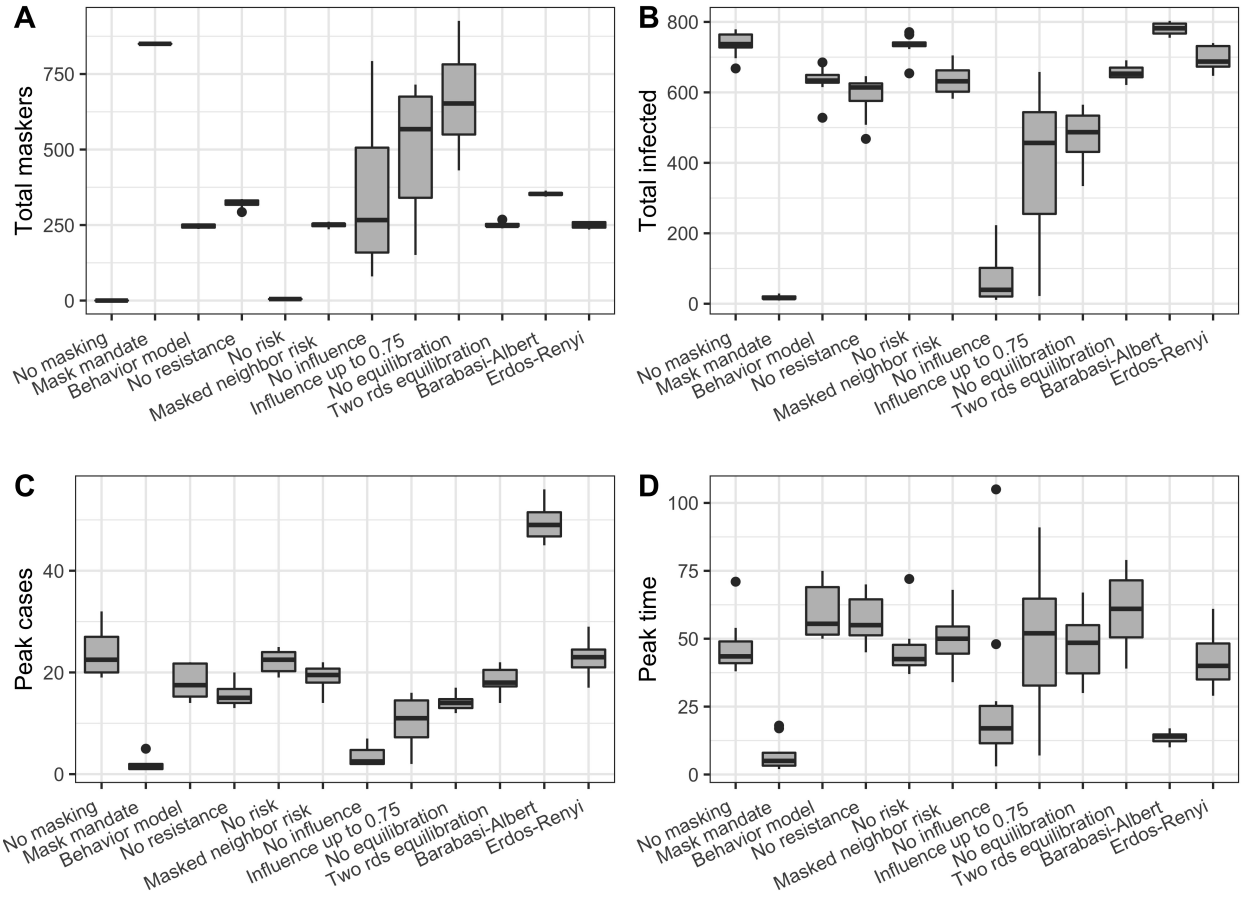

Figure 6.1: Summary of results. Ten simulations conducted on multilayer network 1 with 1000 nodes with varying behavior models. Note that Barabasi-Albert and Erdos-Renyi simulations were conducted on the respective networks. (A) Total number of nodes who masked at some point during the outbreak. (B) Total number of nodes infected. (C) Peak number of daily new infections. (D) Time step of peak number of daily new infections.

6.2 Future Work

We have just scratched the surface in the construction of this coupled disease-behavior dynamical model and there are many possibilities for where we could go next. One option would be to include other non-pharmaceutical intervention behaviors, such as social distancing, quarantine, and isolation. We could take advantage of the option to use dynamic networks to temporarily reduce and restore contacts by removing and creating edges.

Additionally, we could incorporate the perception of global risk of disease into the model. This could be done based on overall disease prevalence in the network, or by evaluating the proportion of infected neighbors one edge, two edges, three edges, or more away from an individual. In an earlier version of our behavioral model we used only symptomatic infectious individuals in the calculation of perceived risk, instead of all infectious individuals. Revisiting this implementation in the context of a community without rapid testing would also be a valuable experiment.

As mentioned earlier, thinking more critically about the scale of influence relative to perceived risk is another component of the model we could further consider. How important should social influence be to decision-making relative to perceived risk? A first step would be to explore the dynamics of more maximum incoming influence values between 0.5 and 1. Moreover, in linear influence games, individuals' thresholds can be negative, which allows individuals to have a predisposition for one behavior over another. Our early experimentation with this idea led people to mask even when there was no disease which was unrealistic. However, this idea of predisposition towards masking, as opposed to just resistance, is one we would like to reflect on further.

Future work may also experiment with different initial configurations of seed nodes or behaviors. For example, we may look to assign similar masking resistances to clusters of individuals so that they behave similarly. Alternatively, we could enforce a mask mandate in some communities of individuals while allowing individuals outside that community to

choose whether or not to mask. Likewise, we could assign a subset of individuals to always mask and another subset to always not mask and then give the remaining individuals a choice on whether or not to mask. These configurations would allow us to further explore the coupled dynamics of masking and disease with greater control over the system.

As more data about COVID-19 becomes available, we would like to incorporate that into the model. Specifically, more information about how infectiousness changes over time could be added by making our β values time-dependent. Times spent in each compartment could also be drawn from a probability distribution that aligned with current data about how long individuals are in the latent and infectious periods.

Finally, we would like to incorporate age structure and more contact heterogeneity into our multilayer network. We could add a school layer that differs from workplace interactions. We could use heavier-tailed degree distributions in each layer to reflect the few individuals that have high numbers of contacts, forming an overall degree distribution between Poisson and power law.

This model serves as a starting point for a variety of future experiments that can help inform predictive disease models and how scientists think about decision-making in an epidemic.

Bibliography

[1] Linda JS Allen, Martha A Jones, and Clyde F Martin, *A discrete-time model with vaccination for a measles epidemic*, Mathematical Biosciences **105** (1991), no. 1, 111–131.

[2] Christian L Althaus, *Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa*, PLoS Currents **6** (2014), ecurrents.outbreaks.91afb5e0f279e7f29e7056095255b288.

[3] Benjamin M Althouse, Edward A Wenger, Joel C Miller, Samuel V Scarpino, Antoine Allard, Laurent Hébert-Dufresne, and Hao Hu, *Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and control*, PLoS Biology **18** (2020), no. 11, e3000897.

[4] Roy M Anderson and Robert M May, *Infectious diseases of humans: dynamics and control*, Oxford University Press, 1992.

[5] David T Arnold, Fergus W Hamilton, Alice Milne, Anna J Morley, Jason Viner, Marie Attwood, Alan Noel, Samuel Gunning, Jessica Hatrick, Sassa Hamilton, et al., *Patient outcomes after hospitalisation with COVID-19 and implications for follow-up: results from a prospective UK cohort*, Thorax **76** (2020), no. 4, 399–401.

[6] Alexander Arnon, John Ricco, and Kent Smetters, *Epidemiological and economic effects of lockdown*, Brookings Papers on Economic Activity (2020).

[7] Paul G Auwaerter, *Coronavirus COVID-19 (SARS-CoV-2)*, Johns Hopkins ABX Guide (2020).

[8] Frank Ball, Denis Mollison, and Gianpaolo Scalia-Tomba, *Epidemics with two levels of mixing*, The Annals of Applied Probability (1997), 46–89.

[9] Shweta Bansal, Bryan T Grenfell, and Lauren Ancel Meyers, *When individual behaviour matters: homogeneous and network models in epidemiology*, Journal of the Royal Society Interface **4** (2007), no. 16, 879–891.

[10] Yinon M Bar-On, Avi Flamholz, Rob Phillips, and Ron Milo, *Science forum: SARS-CoV-2 (COVID-19) by the numbers*, eLife **9** (2020), e57309.

[11] Albert-László Barabási and Réka Albert, *Emergence of scaling in random networks*, Science **286** (1999), no. 5439, 509–512.

[12] Lasko Basnarkov, *SEAIR epidemic spreading model of COVID-19*, Chaos, Solitons & Fractals **142** (2021), 110394.

[13] Chris T Bauch and Samit Bhattacharyya, *Evolutionary game theory and social learning can determine how vaccine scares unfold*, PLoS Computational Biology **8** (2012), no. 4, e1002452.

[14] Chris T Bauch and David JD Earn, *Vaccination and the theory of games*, Proceedings of the National Academy of Sciences **101** (2004), no. 36, 13391–13394.

[15] Chris T Bauch and Alison P Galvani, *Social factors in epidemiology*, Science **342** (2013), no. 6154, 47–49.

[16] Tsanou Berge, Jean M-S Lubuma, G Marcia Moremedi, Neil Morris, and Rose Kondera-Shava, *A simple mathematical model for Ebola in Africa*, Journal of Biological Dynamics **11** (2017), no. 1, 42–74.

[17] Julie C Blackwood, Derek AT Cummings, Hélène Broutin, Sopon Iamsirithaworn, and Pejman Rohani, *Deciphering the impacts of vaccination and immunity on pertussis epidemiology in Thailand*, Proceedings of the National Academy of Sciences **110** (2013), no. 23, 9595–9600.

[18] Anna D Broido and Aaron Clauset, *Scale-free networks are rare*, Nature Communications **10** (2019), no. 1, 1–10.

[19] Victoria J Brookes, Salome Dürr, and Michael P Ward, *Rabies-induced behavioural changes are key to rabies persistence in dog populations: investigation using a network-based model*, PLoS Neglected Tropical Diseases **13** (2019), no. 9, e0007739.

[20] Andrew William Byrne, David McEvoy, Aine B Collins, Kevin Hunt, Miriam Casey, Ann Barber, Francis Butler, John Griffin, Elizabeth A Lane, Conor McAloon, et al., *Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases*, BMJ Open **10** (2020), no. 8, e039856.

[21] Craig Calhoun, *Prisoner's dilemma*, Oxford University Press [Internet]. [cited 07 May 2021]. Available: https://www.oxfordreference.com/view/10.1093/acref/9780195123715.001.0001/acref-9780195123715-e-1338 (2002).

[22] Angelo Carfì, Roberto Bernabei, Francesco Landi, et al., *Persistent symptoms in patients after acute COVID-19*, JAMA **324** (2020), no. 6, 603–605.

[23] Centers for Disease Control and Prevention, *How to protect yourself & others*, [Internet]. [cited 07 May 2021]. Available: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html.

[24] _____, *Clinical questions about COVID-19: questions and answers*, [Internet]. [cited 07 May 2021]. Available: https://www.cdc.gov/coronavirus/2019-ncov/hcp/faq.html (2020).

[25] _____, *COVID-19 pandemic planning scenarios*, [Internet]. [cited 07 May 2021]. Available: https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html (2021).

[26] _____, *Interim guidance on ending isolation and precautions for adults with COVID-19*, [Internet]. [cited 07 May 2021]. Available: https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html (2021).

[27] Muge Cevik, Matthew Tate, Ollie Lloyd, Alberto Enrico Maraolo, Jenna Schafers, and Antonia Ho, *SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis*, The Lancet Microbe **2** (2020), no. 1, e13–e22.

[28] Mustapha Chamekh, Maud Deny, Marta Romano, Nicolas Lefèvre, Francis Corazza, Jean Duchateau, and Georges Casimir, *Differential susceptibility to infectious respiratory diseases between males and females linked to sex-specific innate immune inflammatory response*, Frontiers in Immunology **8** (2017), 1806.

[29] Frederick Chen, Miaohua Jiang, Scott Rabidoux, and Stephen Robinson, *Public avoidance and epidemics: insights from an economic model*, Journal of Theoretical Biology **278** (2011), no. 1, 107–119.

[30] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani, *The role of the airline transportation network in the prediction and predictability of global epidemics*, Proceedings of the National Academy of Sciences **103** (2006), no. 7, 2015–2020.

[31] Kenneth L Cooke and Pauline Van Den Driessche, *Analysis of an SEIRS epidemic model with two delays*, Journal of Mathematical Biology **35** (1996), no. 2, 240–260.

[32] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, *The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2*, Nature Microbiology **5** (2020), no. 4, 536.

[33] Sara Y Del Valle, James M Hyman, and Nakul Chitnis, *Mathematical models of contact patterns between age groups for predicting the spread of infectious diseases*, Mathematical Biosciences and Engineering **10** (2013), 1475.

[34] Ensheng Dong, Hongru Du, and Lauren Gardner, *An interactive web-based dashboard to track COVID-19 in real time*, The Lancet Infectious Diseases **20** (2020), no. 5, 533–534.

[35] David Easley and Jon Kleinberg, *Networks, crowds, and markets*, Cambridge University Press, 2012.

[36] Joshua M Epstein, Jon Parker, Derek Cummings, and Ross A Hammond, *Coupled contagion dynamics of fear and disease: mathematical and computational explorations*, PLoS One **3** (2008), no. 12, e3955.

[37] Dennis M Feehan and Ayesha S Mahmud, *Quantifying population contact patterns in the United States during the COVID-19 pandemic*, Nature Communications **12** (2021), no. 1, 1–9.

[38] Neil Ferguson, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunuba Perez, Gina Cuomo-Dannenburg, et al., *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*, [Internet]. [cited 07 May 2021]. Available: https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/ (2020).

[39] Feng Fu, Daniel I Rosenbloom, Long Wang, and Martin A Nowak, *Imitation dynamics of vaccination behaviour on social networks*, Proceedings of the Royal Society B: Biological Sciences **278** (2011), no. 1702, 42–49.

[40] Eriko Fukuda, Satoshi Kokubo, Jun Tanimoto, Zhen Wang, Aya Hagishima, and Naoki Ikegaya, *Risk assessment for infectious disease and its impact on voluntary vaccination behavior in social networks*, Chaos, Solitons & Fractals **68** (2014), 1–9.

[41] Sebastian Funk, Erez Gilad, Chris Watkins, and Vincent AA Jansen, *The spread of awareness and its impact on epidemic outbreaks*, Proceedings of the National Academy of Sciences **106** (2009), no. 16, 6872–6877.

[42] Clara Granell, Sergio Gómez, and Alex Arenas, *Dynamical interplay between awareness and epidemic spreading in multiplex networks*, Physical Review Letters **111** (2013), no. 12, 128701.

[43] ———, *Competing spreading processes on multiplex networks: awareness and epidemics*, Physical Review E **90** (2014), no. 1, 012808.

[44] Wei-jie Guan, Zheng-yi Ni, Yu Hu, Wen-hua Liang, Chun-quan Ou, Jian-xing He, Lei Liu, Hong Shan, Chun-liang Lei, David SC Hui, et al., *Clinical characteristics of coronavirus disease 2019 in China*, New England Journal of Medicine **382** (2020), no. 18, 1708–1720.

[45] John Guare, *Six degrees of separation*, Performance, 1990.

[46] Sunetra Gupta, Roy M Anderson, and Robert M May, *Networks of sexual contacts: implications for the pattern of spread of HIV*, AIDS (London, England) **3** (1989), no. 12, 807–817.

[47] Qingmei Han, Qingqing Lin, Zuowei Ni, and Liangshun You, *Uncertainties about the transmission routes of 2019 novel coronavirus*, Influenza and Other Respiratory Viruses **14** (2020), no. 4, 470–471.

[48] Craig A Harper, Liam P Satchell, Dean Fido, and Robert D Latzman, *Functional fear predicts public health compliance in the COVID-19 pandemic*, International Journal of Mental Health and Addiction (2020), 1–14.

[49] Xi He, Eric HY Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y Wong, Yujuan Guan, Xinghua Tan, et al., *Temporal dynamics in viral shedding and transmissibility of COVID-19*, Nature Medicine **26** (2020), no. 5, 672–675.

[50] Herbert W Hethcote and James A Yorke, *Gonorrhea transmission dynamics and control*, vol. 56, Springer, 2014.

[51] Inga Holmdahl and Caroline Buckee, *Wrong but useful—what COVID-19 epidemiologic models can and cannot tell us*, New England Journal of Medicine **383** (2020), no. 4, 303–305.

[52] Mevin B Hooten, Jessica Anderson, and Lance A Waller, *Assessing North American influenza dynamics with a statistical SIRS model*, Spatial and Spatio-temporal Epidemiology **1** (2010), no. 2-3, 177–185.

[53] Xun C Huang and Minaya Villasana, *An extension of the Kermack–McKendrick model for AIDS epidemic*, Journal of the Franklin Institute **342** (2005), no. 4, 341–351.

[54] Lars Hufnagel, Dirk Brockmann, and Theo Geisel, *Forecast and control of epidemics in a globalized world*, Proceedings of the National Academy of Sciences **101** (2004), no. 42, 15124–15129.

[55] Mohammad T Irfan and Luis E Ortiz, *On influence, stable behavior, and the most influential individuals in networks: A game-theoretic approach*, Artificial Intelligence **215** (2014), 79–119.

[56] Carl-Joar Karlsson and Julie Rowlett, *Decisions and disease: a mechanism for the evolution of cooperation*, Scientific Reports **10** (2020), no. 1, 1–9.

[57] David Kempe, Jon Kleinberg, and Éva Tardos, *Maximizing the spread of influence through a social network*, Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003, pp. 137–146.

[58] William Ogilvy Kermack and Anderson G McKendrick, *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** (1927), no. 772, 700–721.

[59] Barbara Kersley, Carmen Alpin, John Forth, Alex Bryson, Helen Bewley, Gill Dixon, and Sarah Oxenbridge, *Inside the workplace: first findings from the 2004 workplace employment relations survey*, Department of Trade and Industry, 2005.

[60] Stephen M Kissler, Nishant Kishore, Malavika Prabhu, Dena Goffman, Yaakov Beilin, Ruth Landau, Cynthia Gyamfi-Bannerman, Brian T Bateman, Jon Snyder, Armin S Razavi, et al., *Reductions in commuting mobility correlate with geographic differences*

in SARS-CoV-2 prevalence in New York City, Nature Communications **11** (2020), no. 1, 1–6.

[61] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter, *Multilayer networks*, Journal of Complex Networks **2** (2014), no. 3, 203–271.

[62] Sarah C Kramer, Sen Pei, and Jeffrey Shaman, *Forecasting influenza in Europe using a metapopulation model incorporating cross-border commuting and air travel*, PLoS Computational Biology **16** (2020), no. 10, e1008233.

[63] Kai Kupferschmidt and Jon Cohen, *Will novel virus go pandemic or be contained?*, 2020.

[64] Kathy Leung, Mark Jit, Eric HY Lau, and Joseph T Wu, *Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong*, Scientific Reports **7** (2017), no. 1, 1–12.

[65] Jingjing Li, Yu Liu, TaeHyung Kim, Renqiang Min, and Zhaolei Zhang, *Gene expression variability within and between human populations and implications toward disease susceptibility*, PLoS Computational Biology **6** (2010), no. 8, e1000910.

[66] Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, Jessica Y Wong, et al., *Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia*, New England Journal of Medicine (2020).

[67] Zhirong Liu, Ruilin Chu, Lei Gong, Bin Su, and Jiabing Wu, *The assessment of transmission efficiency and latent infection period in asymptomatic carriers of SARS-CoV-2 infection*, International Journal of Infectious Diseases **99** (2020), 325–327.

[68] James O Lloyd-Smith, Sebastian J Schreiber, P Ekkehard Kopp, and Wayne M Getz, *Superspreading and the effect of individual variation on disease emergence*, Nature **438** (2005), no. 7066, 355–359.

[69] Ira M Longini, Azhar Nizam, Shufu Xu, Kumnuan Ungchusak, Wanna Hanshaoworakul, Derek AT Cummings, and M Elizabeth Halloran, *Containing pandemic influenza at the source*, Science **309** (2005), no. 5737, 1083–1087.

[70] Ira M Longini Jr, Eugene Ackerman, and Lila R Elveback, *An optimization model for influenza A epidemics*, Mathematical Biosciences **38** (1978), no. 1-2, 141–157.

[71] Ira M Longini Jr, M Elizabeth Halloran, Azhar Nizam, Yang Yang, Shufu Xu, Donald S Burke, Derek AT Cummings, and Joshua M Epstein, *Containing a large bioterrorist smallpox attack: a computer simulation approach*, International Journal of Infectious Diseases **11** (2007), no. 2, 98–108.

[72] Shujuan Ma, Jiayue Zhang, Minyan Zeng, Qingping Yun, Wei Guo, Yixiang Zheng, Shi Zhao, Maggie H Wang, and Zuyao Yang, *Epidemiological parameters of coronavirus disease 2019: a pooled analysis of publicly reported individual data of 1155 cases from seven countries*, https://www.medrxiv.org/content/10.1101/2020.03.21.20040329v1 (2020).

[73] Anna Machens, Francesco Gesualdo, Caterina Rizzo, Alberto E Tozzi, Alain Barrat, and Ciro Cattuto, *An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices*, BMC Infectious Diseases **13** (2013), no. 1, 1–15.

[74] Zachary J Madewell, Yang Yang, Ira M Longini, M Elizabeth Halloran, and Natalie E Dean, *Household transmission of SARS-CoV-2: A systematic review and meta-analysis*, JAMA Network Open **3** (2020), no. 12, e2031756–e2031756.

[75] Lauren Ancel Meyers, Babak Pourbohloul, Mark EJ Newman, Danuta M Skowronski, and Robert C Brunham, *Network theory and SARS: predicting outbreak diversity*, Journal of Theoretical Biology **232** (2005), no. 1, 71–81.

[76] Stanley Milgram, *The small world problem*, Psychology Today **2** (1967), no. 1, 60–67.

[77] Dina Mistry, Maria Litvinova, Ana Pastore y Piontti, Matteo Chinazzi, Laura Fumanelli, Marcelo FC Gomes, Syed A Haque, Quan-Hui Liu, Kunpeng Mu, Xinyue Xiong, et al., *Inferring high-resolution human mixing patterns for disease modeling*, Nature Communications **12** (2021), no. 1, 1–12.

[78] Stephen S Morse, Jonna AK Mazet, Mark Woolhouse, Colin R Parrish, Dennis Carroll, William B Karesh, Carlos Zambrana-Torrelio, W Ian Lipkin, and Peter Daszak, *Prediction and prevention of the next pandemic zoonosis*, The Lancet **380** (2012), no. 9857, 1956–1965.

[79] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al., *Social contacts and mixing patterns relevant to the spread of infectious diseases*, PLoS Med **5** (2008), no. 3, e74.

[80] Anjalika Nande, Ben Adlam, Justin Sheen, Michael Z Levy, and Alison L Hill, *Dynamics of COVID-19 under social distancing measures are driven by transmission network structure*, PLoS Computational Biology **17** (2021), no. 2, e1008684.

[81] John Nash, *Non-cooperative games*, Annals of Mathematics (1951), 286–295.

[82] Martial L Ndeffo Mbah, Jingzhou Liu, Chris T Bauch, Yonas I Tekel, Jan Medlock, Lauren Ancel Meyers, and Alison P Galvani, *The impact of imitation on vaccination behavior in social contact networks*, PLoS Computational Biology **8** (2012), no. 4, e1002469.

[83] Mark EJ Newman, *The structure of scientific collaboration networks*, Proceedings of the National Academy of Sciences **98** (2001), no. 2, 404–409.

[84] ———, *Spread of epidemic disease on networks*, Physical Review E **66** (2002), no. 1, 016128.

[85] Cagri Ozcaglar, Amina Shabbeer, Scott L Vandenberg, Bülent Yener, and Kristin P Bennett, *Epidemiological models of mycobacterium tuberculosis complex infections*, Mathematical Biosciences **236** (2012), no. 2, 77–96.

[86] Tiago P. Peixoto, *The graph-tool python library*, figshare (2014).

[87] Piero Poletti, Marco Ajelli, and Stefano Merler, *Risk perception and effectiveness of uncoordinated behavioral responses in an emerging epidemic*, Mathematical Biosciences **238** (2012), no. 2, 80–89.

[88] Travis C Porco, Karen A Holbrook, Susan E Fernyak, Diane L Portnoy, Randy Reiter, and Tomás J Aragón, *Logistics of community smallpox control through contact tracing and ring vaccination: a stochastic network model*, BMC Public Health **4** (2004), no. 1, 1–20.

[89] Benjamin Rader, Samuel V Scarpino, Anjalika Nande, Alison L Hill, Ben Adlam, Robert C Reiner, David M Pigott, Bernardo Gutierrez, Alexander E Zarebski, Munik Shrestha, et al., *Crowding and the shape of COVID-19 epidemics*, Nature Medicine **26** (2020), no. 12, 1829–1834.

[90] Timothy C Reluga, *Game theory of social distancing in response to an epidemic*, PLoS Computational Biology **6** (2010), no. 5, e1000793.

[91] Timothy C Reluga and Alison P Galvani, *A general approach for population games with application to vaccination*, Mathematical Biosciences **230** (2011), no. 2, 67–78.

[92] Steven Riley and Neil M Ferguson, *Smallpox transmission and control: spatial dynamics in Great Britain*, Proceedings of the National Academy of Sciences **103** (2006), no. 33, 12637–12642.

[93] Alessandro Rizzo, Biagio Pedalino, and Maurizio Porfiri, *A network model for Ebola spreading*, Journal of Theoretical Biology **394** (2016), 212–222.

[94] Masaya M Saito, Hiroshi Nishiura, and Tomoyuki Higuchi, *Reconstructing the transmission dynamics of rubella in Japan, 2012-2013*, PloS One **13** (2018), no. 10, e0205889.

[95] Lisa Sattenspiel and Alun Lloyd, *The geographic spread of infectious diseases: models and applications*, vol. 5, Princeton University Press, 2009.

[96] K Sheikh, D Watkins, J Wu, and M Gröndahl, *How bad will the coronavirus outbreak get? Here are 6 key factors*, The New York Times (2020).

[97] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Lorenzo Isella, Corinne Régis, Jean-François Pinton, Nagham Khanafer, Wouter Van den Broeck, et al., *Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees*, BMC Medicine **9** (2011), no. 1, 1–15.

[98] Mark W Tenforde, Sara S Kim, Christopher J Lindsell, Erica Billig Rose, Nathan I Shapiro, D Clark Files, Kevin W Gibbs, Heidi L Erickson, Jay S Steingrub, Howard A Smithline, et al., *Symptom duration and risk factors for delayed return to usual health among outpatients with COVID-19 in a multistate health care systems network—United States, March–June 2020*, Morbidity and Mortality Weekly Report **69** (2020), no. 30, 993.

[99] The United States Census Bureau, *America's families and living arrangements: 2017*, [Internet]. [cited 07 May 2021]. Available: https://www.census.gov/data/tables/2019/demo/families/cps-2019. html (2018).

[100] Robin N Thompson, Christopher A Gilligan, and Nik J Cunniffe, *Detecting presymptomatic infection is necessary to forecast major epidemics in the earliest stages of infectious disease outbreaks*, PLoS Computational Biology **12** (2016), no. 4, e1004836.

[101] Calvin Tsay, Fernando Lejarza, Mark A Stadtherr, and Michael Baldea, *Modeling, state estimation, and optimal control for the us COVID-19 outbreak*, Scientific Reports **10** (2020), no. 1, 1–12.

[102] Brian Uzzi, Luis AN Amaral, and Felix Reed-Tsochas, *Small-world networks and management science research: a review*, European Management Review **4** (2007), no. 2, 77–91.

[103] Raffaele Vardavas, Aaron Strong, Jennifer Bouey, Jonathan William Welburn, Pedro Nascimento de Lima, Lawrence Baker, Keren Zhu, Michelle Priest, Lynn Hu, Jeanne S Ringel, et al., *The health and economic impacts of nonpharmaceutical interventions to address COVID-19*, RAND Report TLA **173** (2020).

[104] Erik M Volz, Joel C Miller, Alison Galvani, and Lauren Ancel Meyers, *Effects of heterogeneous and clustered contact patterns on infectious disease dynamics*, PLoS Computational Biology **7** (2011), no. 6, e1002042.

[105] John Von Neumann and Oskar Morgenstern, *Theory of games and economic behavior (commemorative edition)*, Princeton University Press, 2007.

[106] Zhen Wang, Michael A Andrews, Zhi-Xi Wu, Lin Wang, and Chris T Bauch, *Coupled disease–behavior dynamics on complex networks: a review*, Physics of Life Reviews **15** (2015), 1–29.

[107] Sui-Lee Wee and Donald G McNeil Jr, *China identifies new virus causing pneumonia-like illness*, New York Times (2020).

[108] Wycliffe E Wei, Zongbin Li, Calvin J Chiew, Sarah E Yong, Matthias P Toh, and Vernon J Lee, *Presymptomatic transmission of SARS-CoV-2—Singapore, January 23–March 16, 2020*, Morbidity and Mortality Weekly Report **69** (2020), no. 14, 411.

[109] World Health Organization, *Media briefing on COVID-19 with Dr. Tedros*, [Internet]. [cited 07 May 2021]. Available: https://www.pscp.tv/w/1djxXQkqApVKZ.

[110] _____, *Coronavirus disease 2019 (COVID-19): situation report, 22*, [Internet]. [cited 07 May 2021]. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf?sfvrsn=fb6d49b1_2 (2020).

[111] Shang Xia and Jiming Liu, *A computational approach to characterizing the impact of social influence on individuals' vaccination decision making*, PLoS One **8** (2013), no. 4, e60373.

[112] Hai-Feng Zhang, Zimo Yang, Zhi-Xi Wu, Bing-Hong Wang, and Tao Zhou, *Braess's paradox in epidemic game: better condition results in less payoff*, Scientific Reports **3** (2013), no. 1, 1–8.

[113] Ze-Yu Zhao, Yuan-Zhao Zhu, Jing-Wen Xu, Shi-Xiong Hu, Qing-Qing Hu, Zhao Lei, Jia Rui, Xing-Chun Liu, Yao Wang, Meng Yang, et al., *A five-compartment model of age-specific transmissibility of SARS-CoV-2*, Infectious Diseases of Poverty **9** (2020), no. 1, 1–15.