Infectious Disease Spread on Context-Specific Interaction Networks: Influenza and COVID-19

by

Kristina Mallory

M.Sc., Brown University; Providence, RI, 2015

B.Sc., University of Central Florida; Orlando, FL, 2014

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in The Division of Applied Mathematics at Brown University

PROVIDENCE, RHODE ISLAND

May 2021

This dissertation by Kristina Mallory is accepted in its present form
by The Division of Applied Mathematics as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.

Date_____ _____

Björn Sandstede, Ph.D., Director

Recommended to the Graduate Council

Date_____ _____

Matthew T. Harrison, Ph.D., Reader

Date_____ _____

Martin R. Maxey, Ph.D., Reader

Approved by the Graduate Council

Date_____ _____

Andrew G. Campbell, Dean of the Graduate School

Curriculum Vitae

Education

- **Ph.D. in Applied Mathematics,** 2021

  Brown University, Providence, RI, USA

  - National Science Foundation (NSF) Graduate Research Fellow, 2014

- **M.Sc. in Applied Mathematics,** 2015

  Brown University, Providence, RI, USA

  - Heidelberg Laureate Forum international "Young Researcher," 2015

- **B.Sc. in Mathematics,** with Honors, 2014

  University of Central Florida (UCF), Orlando, FL, USA

  - Summa Cum Laude, Final GPA: 4.0, Highest academic rank in the College of Sciences

Teaching Experience

Instructor:

- **Calculus I,** Summer 2019

  Mathematics Department, Brown University

- **Linear Algebra,** Summer 2018

  Mathematics Department, Brown University

- **Honors: Partial Differential Equations,** Summer 2017

  Applied Mathematics Department, Brown University

- **Honors: Ordinary Differential Equations,** Summer 2016

  Applied Mathematics Department, Brown University

Teaching Assistant:

- **Honors: Partial Differential Equations,** Spring 2016

  Applied Mathematics Department, Brown University

- **Honors: Ordinary Differential Equations,** Fall 2015

  Applied Mathematics Department, Brown University

- **Calculus I,** Fall 2011

  Mathematics Department, University of Central Florida

Pre-College Course Instructor:

- **"Math Class and the Real World,"** Summer 2016, 2017 & 2018

  International Pre-College Program, Brown University

- **"Geometry in Dynamics,"** Summer 2015

  STEM II: Science for Middle School Program, Brown University

Publications

- **K. Mallory** and R.A. Van Gorder. "Stationary solutions for the 1+1 nonlinear Schrödinger equation modeling repulsive Bose-Einstein condensates in small potentials." Physical Review E, **88** (2013) 013205 (doi: 10.1103/PhysRevE.88.013205).

- **K. Mallory** and R.A. Van Gorder. "Control of error in the homotopy analysis of solutions to the Zakharov system with dissipation." Numerical Algorithms, **64** (2013) 633–657 (doi: 10.1007/s11075-012-9683-6).

- **K. Mallory** and R.A. Van Gorder. "Stationary solutions for the 2+1 nonlinear Schrödinger equation modeling Bose-Einstein condensates in radial potentials." Physical Review E, **90** (2014) 023201 (doi: 10.1103/PhysRevE.90.023201).

- **K. Mallory** and R.A. Van Gorder. "Stationary solutions for the 1+1 nonlinear Schrödinger equation modeling attractive Bose-Einstein condensates in small potentials." Physical Review E, **89** (2014) 013204 (doi: 10.1103/PhysRevE.89.013204).

- **K. Mallory**, R.A. Van Gorder, and K. Vajravelu. "Several classes of exact solutions to the 1 + 1 Born-Infeld equation." Communications in Nonlinear Science and Numerical Simulation, **19** (2014) 1669–1674 (doi: 10.1016/j.cnsns.2013.09.034).

- **K. Mallory** and R.A. Van Gorder. "Method for constructing analytical solutions to the Dym initial value problem." Applied Mathematics and Computation, **226** (2014) 67–82 (doi: 10.1016/j.amc.2013.10.019).

- **K. Mallory** and R.A. Van Gorder. "A transformed time-dependent Michaelis–Menten enzymatic reaction model and its asymptotic stability." Journal of

Mathematical Chemistry, **52** (2014) 222–230 (doi: 10.1007/s10910-013-0257-1).

- **K. Mallory** and R.A. Van Gorder. "Optimal homotopy analysis and control of error for solutions to the non-local Whitham equation." Numerical Algorithms, **66** (2014) 843–863 (doi: 10.1007/s11075-013-9765-0).

- R. Li, R.A. Van Gorder, **K. Mallory**, and K. Vajravelu. "Solution method for the transformed time-dependent Michaelis–Menten enzymatic reaction model." Journal of Mathematical Chemistry, **52** (2014) 2494–2506 (doi: 10.1007/s10910-014-0397-y).

- **K. Mallory** and R.A. Van Gorder. "Stationary solutions for the nonlinear Schrödinger equation modeling three-dimensional radial Bose-Einstein condensates in general potentials." Physical Review E, **92** (2015) 013201 (doi: 10.1103/PhysRevE.92.013201).

- **K. Mallory** and R.A. Van Gorder. "Competitive modes for the detection of chaotic parameter regimes in the general chaotic bilinear system of Lorenz type." International Journal of Bifurcation and Chaos, **25** (2015) 1530012 (doi: 10.1142/S0218127415300128).

- R.A. Van Gorder, K. Makowski, **K. Mallory**, and K. Vajravelu. "Self-similar solutions for the nonlinear dispersion of a pollutant into a river flow." Journal of Mathematical Chemistry, **53** (2015) 1523–1536 (doi: 10.1007/s10910-015-0503-9).

- **K. Mallory**, J.R. Abrams, A. Schwartz, M.V. Ciocanel, A. Volkening, and B. Sandstede. "Influenza spread on context-specific networks lifted from interaction-based diary data." Royal Society Open Science, **8** (2021) 191876 (doi: 10.1098/rsos.191876).

Invited Talks

Guest Panelist:

- "Science Policy & Diplomacy: Intergenerational Dialogue," 2017
  AAAS 2017 Annual Meeting, Boston, MA, USA
  - Panel members: J.P. Bourguignon (President of the European Research Council), J. Hopcroft (Turing Award Winner), E. Maskin (Nobel Laureate), G. Parlakgül, and K. Mallory
- "International Cooperation in Science," 2016
  AAAS 2016 Annual Meeting, German Embassy, Washington, D.C., USA
  - Panel members: Y. Alponte, V. Cerf (Turing Award Winner), W. Phillips (Nobel Laureate), and K. Mallory

Guest Speaker:

- Closing Speech for the Heidelberg Laureate Forum, 2015
  Summarizing the week-long workshop, Heidelberg, Germany
  - 1 of 2 honorees (out of 200 postdoctoral and graduate honorees) invited to speak to the Abel, Fields, Nobel, & Turing laureates and the full cohort of honorees

Academic Honors

- Heidelberg Laureate Forum international "Young Researcher," August 2015
  1 of 100 mathematicians honored worldwide for promising future research
  - Selected among graduate, postdoctoral, & early-career researchers by a panel of international mathematics experts to meet and discuss research with Abel, Fields, & Turing laureates for one week in Heidelberg, Germany
- National Science Foundation Graduate Research Fellowship, Spring 2014
  1 of 17 undergraduate & graduate students selected in Applied Math nationwide
- Hernandez Mathematics Award: Excellence in Undergraduate Achievement, Spring 2014
  1 undergraduate selected each year in the UCF Mathematics Department
- Founders' Scholar Award: Excellence in Academic Performance, Spring 2013 & 2014
  Awarded for highest achievements in the UCF Honors College
- Distinguished service: Florida Campus Compact AmeriCorps Program, Spring 2012
  Awarded for creating and teaching courses in 5 elementary, middle, & high schools
- Second-Place Award in UCF Campus-Wide Writing Competition, Spring 2011
  Awarded as a first-year student on the topic of "Sustainability"

Teaching Development

- **Teaching Consultant**, Fall 2016 – Spring 2017
  Sheridan Center for Teaching and Learning, Brown University
  - Conduct teaching observations of peers in all disciplines & give feedback on inclusive teaching, active learning, student engagement, and clarity
- **Seminar Leader for Teaching Certificate I**, Fall 2016
  Sheridan Center for Teaching and Learning, Brown University
  - Lead discussions & hands-on workshops about active learning, assessment strategies, communication skills, course design, and inclusive teaching
- **Earned Teaching Certificate I**, Fall 2015 – Spring 2016
  Sheridan Center for Teaching and Learning, Brown University
  - Topics: active learning, assessment strategies, communication skills, course design, and inclusive teaching

Service

- **Research & Education Consultant**, Fall 2017
  For the HLFF and GSO board of directors, Heidelberg, Germany
  - Sole graduate student consultant among international professors, post-docs, & researchers, invited by the Heidelberg Laureate Forum Foundation (HLFF) and German Scholars Organization e.V. (GSO)
  - Our academic team joined consultants in computer science, politics, and research to develop an interactive online platform where researchers find

collaborators and funding, policy-makers find relevant results and researchers to interpret them, high schools find introductory research and guest speakers, and all parties find practices to increase diversity

- **Chapter President**, Fall 2016 – Summer 2017
  Society for Industrial & Applied Mathematics (SIAM), Brown University
- **Chapter Vice-President**, Fall 2015 – Summer 2016
  Society for Industrial & Applied Mathematics (SIAM), Brown University
- **Chapter President**, Fall 2014 – Summer 2015
  Association of Women in Mathematics (AWM), Brown University
- **Burnett Community Scholar and Instructor**, Fall 2010 – Summer 2012
  400 hours of service in local schools, Orlando, FL
  - Created and taught 8 different courses in reading, mathematics, and environmental science for advanced students and for struggling students in 5 different elementary, middle, and high schools

Mentoring

- **"Properties of three-dimensional surfaces" research project**, Fall 2019
  Undergraduate student earning a mathematics major, Yale University
- **"Population dynamics" research project**, Fall 2018
  Undergraduate student earning a mathematics major, Brown University

Professional Associations

- American Mathematical Society (AMS)
- American Chemical Society (ACS)
- Society of Industrial and Applied Mathematics (SIAM)
- Association for Women in Mathematics (AWM)

For my students at Brown University —

who motivate, energize, and inspire me

more than anyone else.

I am here because of you. And forever after, I am here for you.

Acknowledgments

It is easy to stamp the date on a finished project to remember the moment it became a reality. But this work did not become a reality one moment in May 2021. This work became a reality in a collection of moments that are as responsible for writing these pages as I am. This work became a reality in moments long before.

... with the freedom to personalize my experience

It feels silly to say that the best student leader is the one who lets the student lead. But after your guidance, Björn, I am convinced it is the only way. You encourage your students to discover their own style, explore their own interests, and ask their own questions. Thank you for being open to every new idea, project, or teaching opportunity I presented. You fully prioritized making this work for me. And boy, did you succeed.

... through the impromptu channels of a Zoom call (or four)

Dr. Harrison, Dr. Maxey, Al, and Veronica: our discussions enlivened the pages of this book and will undoubtedly inspire my work long after Brown. Thank you for reminding me how rewarding it is to do math with other people. And – on a personal note – thank you for showing me that good scientific writing demands that

I use conjunctions. You were right. Call me a conjunction convert!

... in a faraway place called Florida

The day I signed up to teach my first class, Dr. Brigham, was the day I began marveling at your ability to hear, value, and include your students. Incidentally, it was also the day I began not-so-secretly adapting your teaching style as my own. Thank you for those pep talks in undergrad and the time spent perfecting my proofs. You made a world of difference for me. But you make a world of difference for everyone. And that's the thing: you never differentiate or favor. You give your all to every student you teach, and that is why I admire you most.

... amidst the ramblings, outbursts, and energetic musings of one idea and the next

Mom and Dad, you were there the day I decided to be a professor at the ripe old stage of Calculus III ...and then you were there for every other plan that followed. You are my sounding board and my think tank, and you don't consider how things look or if I can. Your support depends on my excitement alone. You would be happy whether I earned a Ph.D. or I pursued any other goal that thrilled me instead. And that is the greatest gift I know.

... by the examples of those I admire most

Kurt, thank you for showing me what it means to have a dream and to go for it with all of your heart. If I studied mathematics the way you study your craft — well, I would be a heck of a lot better.

... and in the company of so many friends.

After all, a wise man once told me that mathematics merely defines the outline of our world. But experience and passion and emotion fill it in with color. During my time at Brown, I have met a wealth of beautiful hues: Alex, Brian, Shamil, Pooja, Bobby, Liam, Luke, Matt, P.L.obby, Marissa, Albert, Cassandra, and Laurie. Thank you for painting all the empty spaces with friendship, encouragement, and perspective.

And to my brightest pigments:

Jeffrey. My first math friend and my greatest confidant. What would PDE and tensors have been without your jokes and your shoulder to lean on?

Laura. A constant on this journey and the wisest soul I know. Sometimes I wonder if I could have done it without you.

Johnny. The one who challenges me most, yet believes in me even more. You are my favorite color. And you made the whole picture come together.

P.S. When I was a bright-eyed undergraduate student dreaming of my one-day Ph.D. thesis, I resolved to acknowledge Kurt Cobain. After so many years, I can't even remember why. But a promise is a promise. And so, Kurt Cobain, that day is today.

Contents

Dedication xii

Acknowledgments xiv

1 Introduction 1

2 Preliminaries 8
2.1 Overview of the Warwick Study 9
2.2 Previous Method of Extending the Data 11

3 Context-Specific Algorithms for Lifting the Diary Data 12
3.1 Home Network Construction 14
3.2 Social Network Construction 18
3.3 Work Network Construction 22

4 Influenza Spread on a Representative Network 27
4.1 The Importance of a Context-Specific Network Topology 31
4.2 Interaction Context Has a High Impact on Disease Progression 33
4.3 Dynamic Responses to Disease Substantially Reduce Epidemic Size . 36
4.4 Model Results Capture Trends in NHS Flu Call Data 37

5 A Context-Specific Network for Providence, Rhode Island 41
5.1 Home Network Construction 43
5.2 Social Network Construction 44
5.3 Work Network Construction 46

6 COVID-19 Dynamics in Providence, Rhode Island 49
6.1 Calibrating the Basic Reproduction Number R_0 52
6.2 Work Interactions Maintain Disease Transmission 55

7 Intervention Strategies and the Spread of COVID-19 57
7.1 Lockdown in Providence, RI during the COVID-19 Pandemic 58
7.2 Small-Business Interactions Only Mildly Impact Disease Progression . 63
7.3 Contact Tracing Does Not Reduce the Total Infection Count 67

8 Discussions and Future Work 72

List of Tables

7.1 Weekly interaction changes to reproduce Providence, RI infection data 60
7.2 Weekly interaction changes where small businesses remain open . . . 65
7.3 Weekly interaction changes where small businesses remain open and social interactions decrease partially 66

List of Figures

2.1 Key features of the diary-based data collected by Read *et al.* in (1) . 10

3.1 Home sub-network details 16
3.2 Social sub-network details 20
3.3 Work sub-network details 25

4.1 Performance of our discrete context-specific SIR model over the classical continuous SIR model 32
4.2 Role of interaction context in influenza spread 34
4.3 Impact of dynamic context-specific responses on disease spread 37
4.4 Comparison of our results with influenza data (2) from England. . . . 39

5.1 Home sub-network details for Providence, RI 44
5.2 Social sub-network details for Providence, RI 46
5.3 Work sub-network details for Providence, RI 47

6.1 Calibration of R_0 in our model (6.0.1) using Providence, RI infection data 53
6.2 Role of interaction context in COVID-19 dynamics 56

7.1 Relationship between interaction rates and disease transmission during the lockdown in Providence, RI 61
7.2 Impact of a large-businesses-only closure on the spread of COVID-19 64
7.3 Inefficiency of contact tracing in the case of COVID-19 70

Chapter One

Introduction

The spread of infectious disease is inextricably linked to human interaction patterns. In order to develop effective strategies against future outbreaks, it is essential to understand the intricacies of real-world social encounters. Network theory offers indispensable tools for capturing these details, as reviewed in (3, 4, 5, 6, 7). We can model a community using an interaction network, where network nodes represent individuals and edges represent the interactions between them. Some studies (3, 4, 6, 7) use analytic and computational results to build idealized networks, which attempt to provide minimal models for the complex processes involved in realistic network formation. But several other works (1, 8, 9, 10, 11, 12) endeavor to construct more accurate networks of social encounters using real-world data. These data-driven networks often distinguish encounters by context to explore the impact of realistic social structure on disease dynamics (1, 13). In particular, the network structure in home, work, and social settings is intrinsically different: homes are small, fully clustered and distinct; work networks are made up of establishments of various sizes that sparsely connect households; and social networks are highly connected and, as such, act as bridges between more isolated homes and workplaces.

To obtain context-specific interaction structure in networks, various types of real-world data have been used. For example, Yang *et al.* (8) used census data and a travel log to assign personal characteristics and daily activity patterns to individuals in Eemnes, Netherlands. An interaction network was then specified from this data by assuming individuals interact when they are in the same location at the same time. In a related way, Bian *et al.* (9) relied on census data and a travel log to generate individuals' daily activities and locations, then assigned interactions between all individuals within the same location. In (10), a grid-like network of neighborhoods and work locations was constructed based on United States census data, and individuals were assumed to make contact with one of eight neighbors. Eubank *et al.*

(11) used TRANSIMS, a computational tool that is built on transportation infrastructure and census data, to create a large synthetic population of agents, each with their own personal, location, and activity data. Then, similar to the approaches in (8, 9), interactions were assumed among any individuals with the same location and time traits.

We note that the real-world data used in the above examples (8, 9, 10, 11) and several other instances focus on traits of individuals. These data provide very detailed information about agents in a community (i.e., network nodes) and little to no direct information about interactions between individuals (i.e., edges in a network). Consequently, we must make strong assumptions in the interaction structure on the network, potentially losing intricate details that are vital for understanding disease dynamics.

By contrast, interaction-based data (1) offer an alternative way of developing network structure that takes an edge-based – rather than a node-based – perspective. While details about the demographic features and travel patterns of individuals may be minimal in such studies, interactions between study participants are fully specified. This gives us ample information about the edges between nodes in the associated network and removes the need to make strong, sweeping assumptions, like that all individuals interact whenever they are in the same place and time as in (8, 9, 10, 11). Interaction-based data exist in several settings; for example, control policies for sexually transmitted diseases are often based on recorded patterns of sexual encounters (14, 15, 16). However, interaction logs for casual contacts, which are often more frequent and unpredictable, are difficult to acquire and minimally exist in the literature (17).

Yet, for the first time, Read *et al.* (1) conducted a longitudinal survey of real-

world casual contact interactions. Students and staff in a university setting were asked to record all of the individuals with whom they interacted, as well as the context (e.g. home, social, work/school, etc.) in which these interactions took place. This type of diary-based study provides rich information on the interactions of study participants, a feature that is often missing from data-driven networks. Nevertheless, generating a network directly from the interaction-based data (1) is infeasible. Foremost, interactions are only recorded in a small subset of the population (namely, those participating in the survey), so networks that are specified entirely from survey data are limited in size (18). Moreover, the interaction data, provided from the sole perspective of survey participants, is necessarily incomplete. While we have a clear log of the individuals with whom the survey participants interact, we have no information on how those other individuals interact. Thus we must consider strategies for extending and completing the data carefully to produce networks which maintain the interaction structures present in the data.

To that end, Read *et al.* (1) presented a method for completing their interaction data which focused exclusively on the network edges. The authors built sub-networks for each type of interaction from their context-specific data, however, no information was generated to differentiate between nodes. Instead, the nodes were connected randomly in order to maintain the layout of edges recorded in the data. This strategy produced extended networks which did not preserve many context-specific features. As an example, the home sub-network in (1) does not consist of distinct home clusters as one would expect.

Outline of our work

Our work is divided into two main parts. In the first part, we present a new

approach for completing the interaction data (1) which preserves the key features present in the data, but also exhibits a more realistic context-specific structure. We develop sub-networks to represent the encounters in home, social, and work settings separately. To capture realistic features in each sub-network, we generate some information on the nodes in addition to the edge-based information provided by the interaction data. One of our main contributions, therefore, is to address some of the challenges associated with the limited and incomplete nature of diary-based interaction data in a context-specific way. This strategy more faithfully represents features of the real data (1) and allows us to investigate how dynamic responses in each interaction context influence epidemic evolution. In particular, we explore the roles that different interaction contexts play on influenza transmission by simulating a susceptible–infected–recovered (SIR) model of disease spread on our extended diary-based networks. Then, we investigate the impact of dynamic changes to our network, representing human responses to infection, and suggest strategies for reducing the spread of disease. Our results, which agree with previous studies (8, 9), suggest the following:

- disease spreads most frequently at work but remains localized unless other interactions are present (§4.2);
- social interactions are responsible for the wider spreading of an epidemic (§4.2);
- dynamic responses to infection can substantially reduce epidemic size (§4.3);
- individuals with influenza are most likely to reduce their work and social interactions (§4.4); and
- staying home from work or school has the strongest impact on reducing the severity of an influenza outbreak (§4.4).

Additionally, in §4.1, we compare discrete SIR dynamics in the network setting

to the classical ordinary differential equation framework. This analysis validates our context-specific modeling approach, and we find qualitative agreement between disease dynamics on our extended network and data from mild to moderate seasons of influenza outbreak (§4.4). We note that the above results, presented in Chapters 2–4, were published previously in (19).

In the second part, we adapt our context-specific modeling approach to build a larger, real-world network for Providence, RI which captures many of the realistic context-specific features recorded in the interaction data (1). This network separates similarly into a home, social, and work sub-network, allowing us to investigate how dynamic responses in each interaction context influence disease progression. Using an SIR model on the network for Providence, RI, we study the impact of several intervention strategies on the spread of COVID-19. Ultimately, we compare the efficacy of these strategies and suggest scenarios where one strategy may be preferred to another. As an overview, our results indicate the following:

- interactions were substantially reduced in Providence, RI during the COVID-19 pandemic, particularly those in work settings (§7.1);
- a decrease in work interactions is necessary to reproduce the early infection count recorded in Providence, RI (§7.1);
- closing small businesses does not significantly alter the spread of disease (§7.2);
- interactions in large businesses greatly facilitate the epidemic (§7.2); and
- contact tracing is not a feasible method for lowering the number of people infected during a COVID-19 outbreak (§7.3).

Furthermore, we use a discrete SIR model on our network to quantify the decreased interaction rate which reproduces the infection data recorded in Providence, RI at

the start of the COVID-19 pandemic (§7.1). The results of part two are discussed in Chapters 5–7.

We complete our study by exploring the implications of our findings and suggesting directions for future work in Chapter 8.

Chapter Two

Preliminaries

We begin by outlining the key features of the diary-based data (1), herein referred to as the **Warwick data**, which serves as the basis for our work. These features are further summarized in Figure 2.1. Then, we discuss the existing method used by the authors in (1) to extend the Warwick data to a completed network.

2.1 Overview of the Warwick Study

The Warwick data, presented in (1), is a record of the person-to-person interactions of 49 volunteer participants over the course of 14 non-consecutive days. The volunteers, consisting of students and staff at the University of Warwick in Coventry, UK, were asked to keep a log of their conversational interactions on the specified days. These encounters were categorized based on proximity (casual or skin-to-skin) and context (home, social, work/school, shop, travel, or other). The resulting record contains a total of 8661 encounters among 3529 people, which includes the survey participants and each of the individuals with whom they interacted. Following the terminology in (1), we will refer to the survey participants as *egos* and the secondary individuals they encountered as *alters*.

There are two main quantitative measurements obtained from the Warwick data that we rely on to construct our networks for each interaction context. First, for each of the 49 egos, we know the **degree** (total number of unique individuals encountered over the 14 days) in each interaction context:

$$\text{degree of individual } i \text{ in interaction context } c = \sum_{j=1, j\neq i}^{N} a_{ij}^{c},$$

where N is the number of nodes in the network (3529 in the case of the Warwick

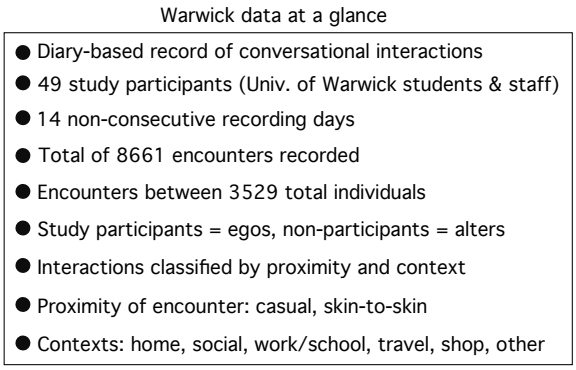

Figure 2.1: Key features of the diary-based data collected by Read *et al.* in (1).

data), and the values of the adjacency matrix A are given by $a_{ij}^c = 1$ if individuals i and j interacted in context c on at least one of the survey days (that is, if i and j are neighbors in the network) and 0 otherwise (see, for instance, (3)). Note that encounters are categorized by interaction context to obtain separate home, social, work, shop, and travel degrees for each participant (thus, for example, two nodes may be neighbors at work, but not at home). These degree measurements are then used to calculate a separate degree distribution (fraction of nodes in the network with degree n (18)) for each interaction context.

Second, the survey data includes a record of repeat interactions over the 14 sample days, and this provides a measure of the strength, or **frequency** $1/14 \leq f_{ij}^c \leq 14/14$, of encounters between individuals i and j in each context c. We note that, because the Warwick data is recorded exclusively from the perspective of the 49 egos, it is challenging to obtain accurate measurements of clustering (a widely-studied quantity related to the connectedness of a graph (20)). For this reason, we focus mainly on degree and frequency distributions. In Chapter 3, we highlight key features of the Warwick data in each context and present our algorithms for extending this structure to larger networks with similar characteristics.

2.2 Previous Method of Extending the Data

The Warwick data (1) offers insight into the nature of real-world interactions as recorded by the survey participants. However, the resulting interaction network is limited by its small sample size. To study disease dynamics on this network, Read *et al.* (1) extended the diary-based Warwick data to a completed network by first making multiple copies of the survey participants, each with the same degree n as an original ego. These copies were represented as a node in the extended network with n unconnected edges or stubs emanating from it. No other defining characteristics were assigned to the nodes in the network. Then, network formation occurred by randomly connecting stubs with the same weight to create edges between nodes. Because this approach is not context-specific, key features that distinguish the structure in home, social, and work/school settings were lost. For example, the extended home network that results from this approach is likely to be highly connected; yet real-world home networks are highly clustered and disconnected.

In Chapter 3, we present our context-specific algorithms for generating an artificial network which accounts for essential differences in home, social, and work/school settings. We elucidate the key features of the Warwick data (1) in each context and assign associated traits to the nodes before connecting edges in the network. This method preserves the interaction structure recorded in the diary-based data (1), while also approximating a network which could be found in the real world.

Chapter Three

Context-Specific Algorithms for Lifting the Diary Data

In this chapter, we present an algorithm for generating an artificial network which maintains some context-specific features of the interaction structure that was recorded in the diary-based data (1). Since home, social, work/school, shop, and travel settings are characteristically very different, our network model consists of context-specific sub-networks. Including shop and travel interactions did not have a strong impact on our results, as discussed in Chapter 8, so we chose to neglect these contexts. Hence we build an extended network in which each individual (or node) can interact with any other individual in one or more of three settings: home, social, and work/school. We represent each interaction context by a separate sub-network on the same nodes; taken together, these sub-networks for home, social, and work/school interactions make up our full extended network.

Since our goal in this chapter is to build on the work of Read *et al.* (1) in a context-specific way, we construct a hypothetical population of 3000 individuals to mimic the individuals interacting in the Warwick data. This data was collected in a small study at the University of Warwick, so we do not expect the network extension to fully capture the interactions of a larger community, such as a college or a city population. It should also be noted that we do not differentiate between proximity of interaction (casual or skin-to-skin). Because this simplification limits the types of diseases we can reliably simulate to those that do not require skin-to-skin contact for transmission, we will focus on influenza and coronavirus dynamics (see Chapters 4, 6, and 7).

Sub-network construction in each setting proceeds in two main steps. First, we build the core of the network, assigning edges between nodes to form a collection of disconnected "units." This step is context-specific: our home network consists of households or student dorms in which individuals live; our social network is based on friend groups; and our work/school network is made up of companies or classrooms.

The size of these units, therefore, differs between contexts (with work units containing more nodes on average than home units, for example). Second, we assign a weight to each edge that reflects the frequency of interaction between the two nodes in that context. We do this by sampling from the respective frequency distribution generated from the Warwick data (1). For example, to assign a weight to a given edge in the home sub-network, we sample a weight from the frequency distribution for home interactions in the Warwick data. These weights, or frequency values f_{ij}^c, take values between 1/14 and 14/14, since the data (1) was collected on 14 days. Network construction for the social sub-network has an additional step where some of the disconnected social units (viz. friend groups) are connected by assigning new edges between the high-degree nodes in each social unit. This accounts for "popular" individuals who interact with multiple friend groups. Connections are then further shuffled in the social network in order to achieve a realistic clustering coefficient seen in online communities (20, 21).

We provide a detailed summary of our network construction for the home, social, and work/school contexts in §3.1, §3.2, and §3.3, respectively.

3.1 Home Network Construction

Our home network is expanded from the Warwick data by adding one home unit at a time. The size of each new household is determined by the degree distribution for Warwick home encounters (1), and we assume that individuals interact with all members of their share household. In particular, suppose we sample from this distribution to obtain a target degree n. We then generate a clique of size $n + 1$, so that each individual has degree n. We also define the average local clustering

coefficient C as

$$C = \frac{1}{N}\frac{\text{number of triangles connected to node } i}{\text{number of triples centered on node } i} = \frac{1}{N}\sum_{i=1}^{N}\frac{2\ell_i}{n_i(n_i-1)},$$

where N is the number of nodes in the network, a triangle is a set of three mutually connected individuals, a triple centered on node i consists of node i and any two nodes it is connected to (whether they are connected to each other or not), n_i is the degree of node i, and ℓ_i is the number of edges in G_i (the sub-graph of neighbors of node i) (18, 20). Note that $C = 1$ for all the nodes in our home network since each home unit is fully connected. Therefore, our approach captures the highly clustered, disconnected structure of the home network. In contrast, because Read *et al.* did not use a context-specific algorithm, the extended home network in (1) is unrealistically connected and does not reproduce this natural clustering feature. We show example home units in Figure 3.1a.

Once we generate a home unit of size $n + 1$, as discussed above, the next step is to determine the frequency, or weight, of each interaction in the household. Because survey participants consisted of students and staff at the University of Warwick, many of these individuals lived in residence halls or shared houses, and this led to high degree and low frequency interactions in the home setting. Students in residence halls may encounter many individuals living in their building, but not necessarily see every housemate every day. In contrast, participants living in residential properties shared by a smaller group of people might be more likely to display frequent interactions with a core group of fewer people at home. As we show in Figure 3.1c, having a high degree is indeed associated with lower average interaction frequency in the Warwick home data. To account for this feature, we separate the frequency distribution from (1) into two components: the distribution for survey participants with home degree less than or equal to 8, and the corresponding distribution for those with degree

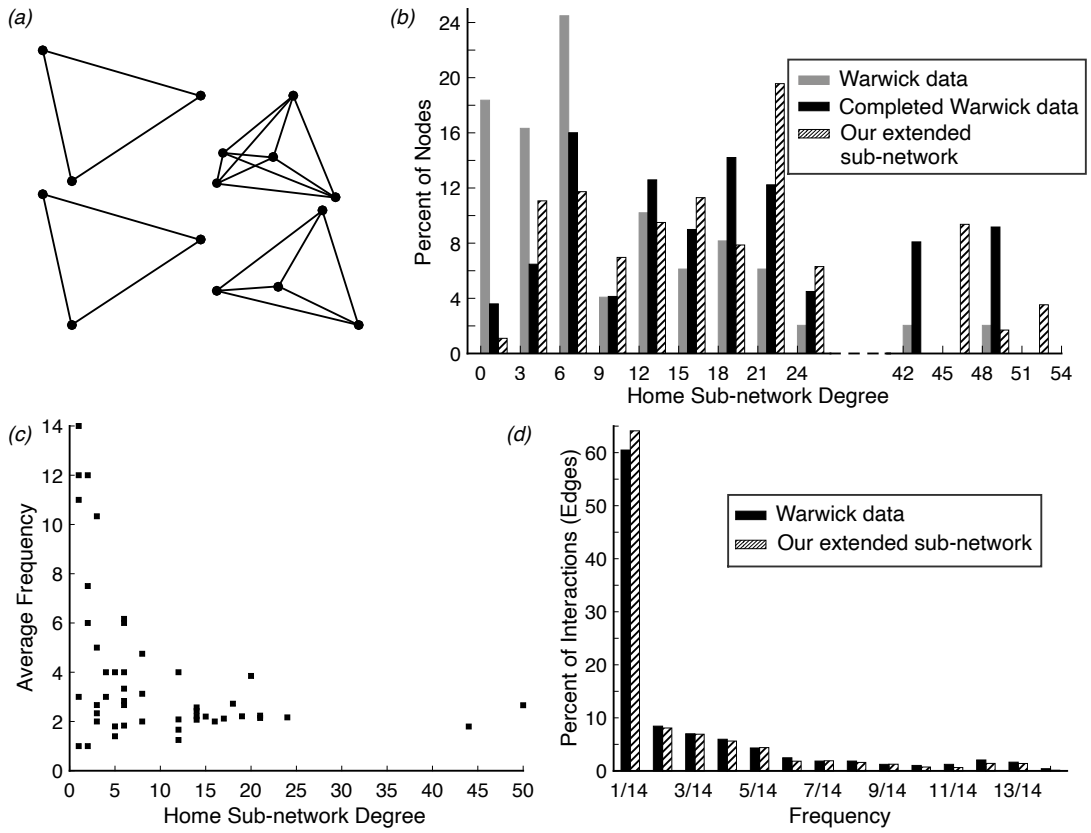

Figure 3.1: Home sub-network. (a) Fully-clustered household units make up our home sub-network; we show 4 example units. (b) The size of each home is determined by sampling from the home degree distribution associated with the Warwick data (1). We also plot the completed Warwick data for direct comparison with our generated network (the Warwick data is completed by assuming each degree n corresponds to a fully clustered household of size $n+1$). (c) Individuals with high degree (inhabitants of large households) interact with other home members less frequently on average, while members of small households encounter a smaller core group more frequently. (d) The distributions for interaction frequencies show good agreement between our algorithm and the Warwick data (1). We plot the raw Warwick data in (b); in (c–d) we plot the Warwick data after accounting for a small number of inconsistencies, such as interactions that were logged by participants on dates that fall outside of the survey days.

greater than 8. Our motivation to use degree 8 to separate these two types of homes comes from our observations of the Warwick data in Figure 3.1c: we see that individuals with degree less than or equal to 8 interact with higher frequencies on average than individuals with degree greater than 8. This value is also similar to the findings of Read *et al.* (1): shared homes had a median household size of less than

7 and residence halls had a median household size of 7.

For small households, we assign weights to each edge by sampling from the frequency distribution for home interactions with degree less than or equal to 8; and, for large households, we assign weights to each edge by sampling from the distribution for degree greater than 8. This completes the home unit. The result is a new, fully-clustered household of size $n + 1$ added to the network, where the degree n of each household member was determined from the diary-based data (1), the frequency of interactions was sampled from the real frequency distribution (1), and the clustering coefficient is 1 for every node, since we assume all-to-all coupling.

We summarize our full home sub-network algorithm below:

1. Sample from the home degree distribution of the Warwick data (1) to obtain a target degree n.
2. Generate a clique of size $n + 1$.
3. Assign an interaction weight to each edge in the home unit by sampling from the appropriate home frequency distribution of the Warwick data (1): the distribution for $n \leq 8$ or for $n > 8$.
4. Add the new home unit to the existing network and repeat from Step 1 until the desired number of nodes in the network are generated (i.e., 3000 nodes in this case). Note we may have to adjust the size of the final home unit to reach the desired number.

We plot the degree and frequency distributions obtained from this algorithm alongside the Warwick data (1) in Figures 3.1b and 3.1d. For comparison, we also plot what we consider to be the *completed* Warwick data in Figure 3.1b. In par-

ticular, because we assume all-to-all coupling within each home, we expect that a Warwick-survey participant (1) with home degree n lives with n other individuals who also have degree n. Such a home, therefore, contains $n+1$ individuals with degree n. Accounting for these extra individuals, who we assume are not survey participants, allows us to get a more accurate degree distribution for the total number of individuals living in homes with the 49 Warwick-survey participants. Thus, we multiply the total number of participants h_n who are recorded as having degree n by $n+1$ to obtain the total number of individuals with degree n. We refer to this new data set as the completed Warwick data and plot its degree distribution in Figure 3.1b. Note that it is possible some of the 49 survey participants resided in the same home. However, we assume that the participants live in separate homes when constructing our network since the survey is small.

3.2 Social Network Construction

In contrast to interactions in home settings, social encounters are much more widespread, and we seek to capture this more connected, less clustered character in our extended social sub-network. This means that we cannot build our social sub-network by specifying all-to-all coupling within social units as we did for our home sub-network, which makes realistic network extension more challenging. To address this, we base our social sub-network construction on a simplified distribution that captures features of the Warwick data (1). As illustrated in Figure 3.2b, the social degree distribution appears to be roughly uniform until degree 36, with a few outliers who have many friends. This observation underlies the construction of our extended social sub-network.

We generate our social sub-network in three steps: first, we construct social units (or friend groups) of 38 nodes each. Within each identical social unit, we assign the nodes a degree from 0 to 36 in a uniform manner to account for the roughly uniform distribution on Warwick social degrees less than or equal to 36. (An inductive argument shows that there is a unique way of specifying a uniform distribution on 38 nodes, and it necessarily forces degree 18 to appear twice.) Next, to capture the appearance of social outliers with many friends, we randomly select popular nodes and connect them to high-degree individuals in other social units. Lastly, we shuffle some connections between nodes to bring the average local clustering coefficient down to values reported in (20, 21), in analogy with the small-world model of Watts and Strogatz (22).

We assign interaction weights to each edge in the social sub-network by again separating the measured frequency distribution (1) for social interactions in the Warwick data into two distributions. As in the case of home networks, we notice that individuals with high degree appear to have fewer repeat interactions on average. Thus, we separate the distribution into one for nodes with degree less than or equal to 18 and another for degree greater than 18. Sampling our frequencies from these distinct distributions generates an extended network whose frequency distribution shows good agreement with the Warwick data (see Figure 3.2c).

Figures 3.2c and 3.2d compare the degree and frequency distributions in our extended social sub-network to the Warwick data (1). We summarize the full social sub-network algorithm below:

1. Generate 79 identical social building blocks of size 38 so that the degree distribution within each social unit is approximately uniform from 0 to 36. (Degree 18 necessarily appears twice.)

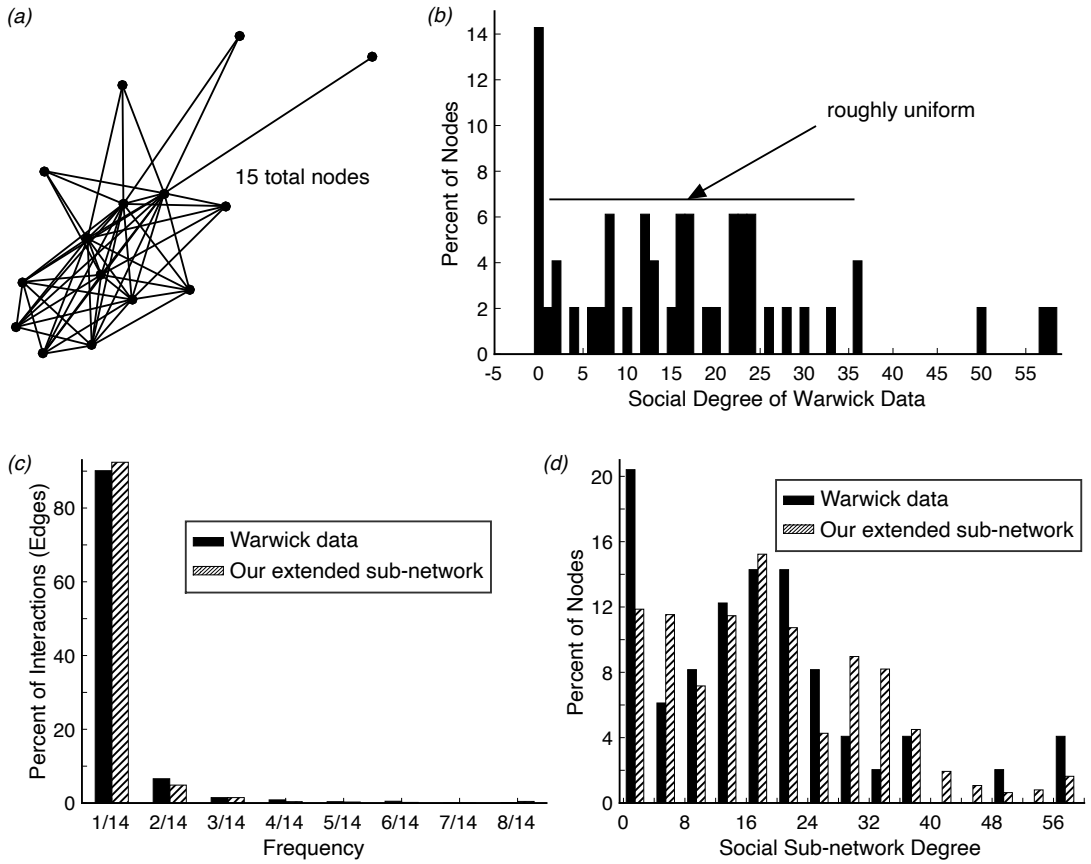

Figure 3.2: Social sub-network. (a) Social units, each with a uniform degree distribution, serve as the basic building block of our extended social sub-network. We show an example network of 15 nodes with a uniform distribution (our social units are each made up of 38 nodes, but a smaller network illustrates the structure more clearly). (b) The social degree distribution for the Warwick data (1) is approximately uniform from degree 0 to 36. Comparison of (c) social frequency and (d) degree distributions for the Warwick data (1) and our extended social sub-network. We plot the raw Warwick data in (b) and (d); in (c) we plot the Warwick data after accounting for a small number of inconsistencies, such as interactions that were logged by participants on dates that fall outside of the survey days.

2. For each social unit, choose α of the β most popular (highest degree) nodes in the unit. Connect each of the chosen nodes to γ high-degree nodes randomly selected from other social units, where "high-degree" means one of the top β highest degree nodes in a social unit.

3. Randomly select M edges and, for each such edge, disconnect one end and reconnect it to another node chosen at random.

4. Randomly select and remove 2 nodes to reduce the total network size to the target 3000 nodes.

5. Assign interaction weights to each edge by sampling from the appropriate social frequency distribution for the Warwick data (1): the distribution for $n \leq 18$ or for $n > 18$.

We use $\alpha = 1$, $\beta = 5$, and $\gamma = 25$. We tested multiple values for these parameters and selected those that best fit the degree distribution of the Warwick data (1). Similarly, we use $M = 13000$ edges, because this reproduces the average local clustering coefficient $C = 0.16$ reported by Ahn *et al.* (21) for the Cyworld social network. (Note that we rely on reports (20, 21) of clustering in online communities to inform our network algorithm because it is difficult to formulate an accurate measure of clustering from (1).) We found that the fewer edges we randomly shuffled, the larger the clustering coefficient, with $C = 0.8$ for $M = 0$ edges. Thus, tuning this parameter could allow for the generation of a network with any given clustering coefficient.

In conclusion, we construct our extended social sub-network in three steps to take into account the key features of the Warwick data and measurements (20, 21) of clustering in online social communities: the first step, building social units of uniform degree, captures the approximately uniform character of the Warwick social degree distribution. The second step, adding edges between randomly selected high-degree members of social units, accounts for high-degree outliers in the Warwick data (1). Lastly, reducing the level of structure in the network by breaking and reconnecting some edges at random brings the clustering coefficient down in agreement with empirically measured values (21). It is also worth noting that reconnecting edges when extending the social sub-network provides connections between more clustered

home and work units in our full network. This step contributes to a fully connected network with a realistically small characteristic path length (22); in particular, we find that the characteristic path length for our full network is 2.71.

3.3 Work Network Construction

We build the work sub-network by constructing work units (companies/schools) as we did for the home and social sub-networks. In this case, however, the work interactions recorded by survey participants in the Warwick data (1) likely occurred within a single work unit, namely, the University of Warwick. Thus we do not directly use the degree distribution for work interactions in the Warwick data (1) to determine the sizes of our work units. Instead, we use data on business sizes in Coventry, UK, as well as qualitative features that we expect in a work environment, to determine appropriate work-unit sizes and assign the edges within them. Then, we assign weights to each edge by sampling from the frequency distribution for work interactions in the Warwick data (1), thereby only using the Warwick data to inform interactions within each unit.

To begin, Read *et al.* (1) found that the degree distribution for casual contacts across all contexts had a significantly longer right tail than the corresponding distribution for skin-to-skin encounters. Since the majority (95.97% (1)) of encounters at work were casual, while most skin-to-skin interactions took place in home or social contexts, we expect that the work/school degree distribution (1) should display a longer right tail than the distributions for other contexts. Indeed, as we show in Figure 3.3a, the Warwick degree distribution for work/school encounters displays a long right-tail, a feature that is characteristic of power-law distributions and often

captured using network-growth models involving preferential attachment (23, 24).

Preferential attachment is a common means of generating networks with scale-free power-law distributions, and it was popularized by the work of Barabási and Albert (24). According to the Barabási-Albert model, network growth occurs by starting with an initial network of m_0 nodes and then adding one node at a time. At each step, the new node is connected to $m \leq m_0$ other nodes, with the probability of connecting to node i given by

$$\Pi(k_i) = \frac{k_i}{\sum_{j=1}^{N} k_j}, \tag{3.3.1}$$

where k_i is the degree of the ith node and N is the total number of nodes in the network. This rule means that new nodes are most likely to connect to existing nodes of high degree, and the result is a network structure in which many nodes are connected to a few very popular (high-degree) individuals.

We build our extended work/school sub-network using preferential attachment motivated by the structure of the work environment (one can think of a business scenario in which many employees interact with a common manager). Alternatively, in the school context, we would envision many students conversing with a few teaching assistants, and everyone interacting with a single course instructor. Therefore, as we add nodes to the network, we want each individual to be more likely to connect to the instructor (node of high degree) than to any given student. The long right tail of the real work/school degree distribution measured from the Warwick data further supports our choice to base network extension on preferential attachment. Although preferential attachment produces distributions with long right tails, it is important to note that this does not necessarily mean networks with long right tails emerge from preferential-attachment dynamics. Our focus is on lifting interaction-based

data to larger networks that maintain the same features, however, and preferential-attachment methods accomplish this goal.

When developing our model, we tested three different implementations of network growth using the idea of preferential attachment. We began by building complete networks one node at a time according to the Barabási-Albert model (24), but this led to networks with degree distributions that were too narrow. To remedy this issue, we also tried a variation of the Barabási-Albert model (24) that was motivated by the fitness model of Bianconi and Barabási (25). In particular, to penalize high-degree nodes from receiving additional edges after they reach a certain degree k_0, we modified the original probability in equation (3.3.1) to the following:

$$\Pi(k_i) = \frac{k_i \eta(k_i)}{\sum_{j=1}^{N} k_j}, \tag{3.3.2}$$

where $\eta(k) = \frac{1}{2}(1 - \tanh \frac{k-k_0}{\varepsilon})$ is a cutoff function. Here we select $\varepsilon, k_0 > 0$ to best fit the real data (1). It should be noted that we have replaced η_i, a native fitness value for each node i that is chosen from a specified distribution in (25), with $\eta(k)$. While the original Barabási-Albert algorithm (24) and our altered version of the fitness model (25) are able to produce degree distributions with the observed long right tail, neither method captures the high amount of clustering reported in the real data (1).

To raise the clustering coefficient, we return to the idea of building networks out of units. Using data on business sizes in the city of Coventry, UK from the Inter Departmental Business Register (IDBR) (26), we construct unit sizes in a 3000-person network to match the real size distribution in Coventry. We begin with the larger businesses, constructing two companies of 310 individuals, three companies of 200 people, five companies of 85 people, and twelve companies of 38 people. We then

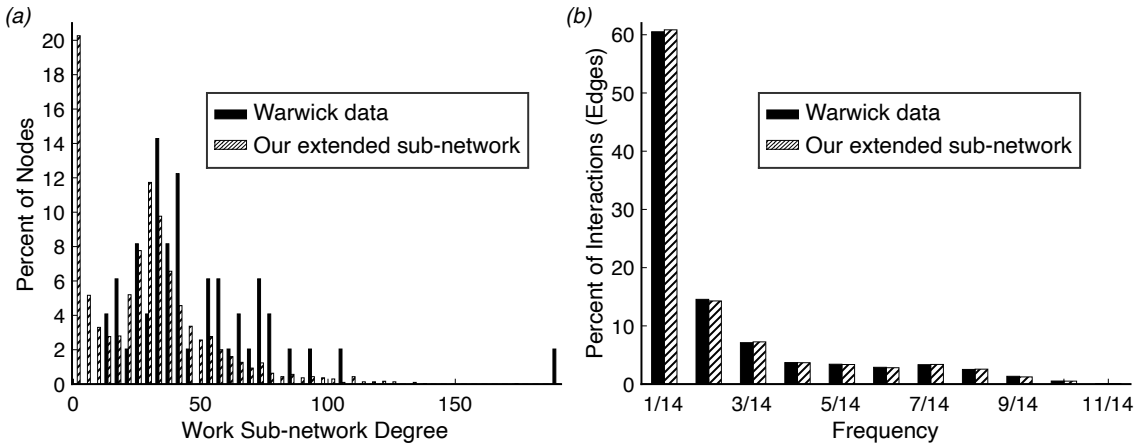

Figure 3.3: Work sub-network. (a) Degree and (b) frequency distributions for the work sub-network show good agreement between the Warwick data (1) and our extended model. We plot the raw Warwick data in (a); in (b) we plot the Warwick data after accounting for some inconsistencies, such as interactions that were logged by participants on dates that fall outside of the survey days.

use the original Barabási-Albert algorithm (24) to generate the degree distribution within each large business. Next, we construct the smaller businesses which have 20 or fewer employees and specify a roughly uniform distribution within each unit. Our choice to use a uniform degree distribution within the small businesses/classrooms is based on the idea that small settings allow for a more interactive structure than larger ones. Additionally, incorporating small work units of uniform degree into the network serves to increase the amount of clustering.

We summarize the details of our full work sub-network algorithm below:

1. Use the original Barabási-Albert model (24) with $m = m_0 = 30$ to generate the following large work units:

   - 2 work units made up of 310 nodes
   - 3 work units made up of 200 nodes
   - 5 work units made up of 85 nodes
   - 12 work units made up of 38 nodes

(We choose this value of m to best match the degree distribution of the real data (1).)

2. Generate the following small work units with a uniform degree distribution in each:

   - 17 work units made up of 17 nodes, each with a degree distribution that is roughly uniform from 0 to 15 (degree 8 will appear twice)
   - 26 work units made up of 8 nodes, each with a degree distribution that is roughly uniform from 0 to 6 (degree 3 will appear twice)
   - 134 work units made up of 3 nodes, each with a degree distribution that is roughly uniform from 0 to 1 (degree 1 will appear twice)

3. Together the work/school units generated in Steps 1–2 represent the network. Assign an interaction weight to each edge in the network by sampling from the work frequency distribution for the Warwick data (1).

By combining the concept of preferential attachment in large businesses (or classrooms) with uniform degree distributions in small work units, we are able to generate an extended work/school sub-network with degree and frequency distributions that capture many of the features of the Warwick data (1), as we show in Figure 3.3.

Chapter Four

Influenza Spread on a Representative Network

We now turn to a study of an epidemic spreading on the extended network that we generate from the Warwick data (1) as described in Chapter 3. We assume a disease that gives long-term immunity after recovery, so that the susceptible–infected–recovered (SIR) model framework is appropriate (27). Individuals can therefore be susceptible (S), infected (I), or recovered (R), and the infected individuals are assumed to be capable of infecting other susceptible agents. We denote the i^{th} susceptible node by S_i and the j^{th} infected node by I_j. We assume each infected node recovers from the disease after a time drawn from an exponential distribution with mean T, where T is the average duration of infectiousness.

We can define the basic reproduction number R_0, a quantity commonly studied for infectious diseases, on our network. Simply put, the basic reproduction number R_0 is the average number of secondary infections produced from one infected individual during the infectious period T in a population that is otherwise fully susceptible. We begin by calculating the total number of susceptible individuals that will be infected by a given infected individual I_j in our network during the infectious period T. We call this number $S_T(j)$. In general, the probability that any infected individual in the network will infect a susceptible individual per unit contact time is the **transmission risk** τ and can be written

$$\tau \coloneqq P\left(\frac{\text{infected individual infects a susceptible individual}}{\text{unit contact time}}\right).$$

Then, the probability of transmission across a particular edge will be proportional to the transmission risk τ and scaled by the edge weight (1). For example, consider a simple network consisting of one infected individual I at time t and one susceptible individual S that interact with a frequency f that is defined to be the

$$f \coloneqq \frac{\text{contact time}}{\text{unit time}}$$

between I and S. Then, the probability that I infects S on the time interval dt is

$$P(I \text{ infects } S \text{ in } [t, t+dt)) \;=\; \tau \cdot \frac{\text{contact time}}{\text{unit time}} \cdot dt \;=\; \tau f \, dt \,.$$

On our extended network, each infected individual I_j can infect any of its susceptible neighbors in each of the three interaction contexts c (i.e., home, social, and work). Let $N^c(j)$ be the set of indices i such that the ith node is connected to I_j in context c. Then, in the scenario where all other nodes in the network are susceptible, I_j will infect

$$S_T(j) \;=\; \sum_{\text{contexts } c} \; \sum_{i \in N^c(j)} \tau \, f_{ij}^c \, T$$

nodes during the infectious period T, where f_{ij}^c is the frequency of interaction between I_j and susceptible node S_i, $i \in N^c(j)$, in context c. Finally, to obtain the basic reproduction number R_0, we take the average of $S_T(j)$ for any such infected individual I_j in the network. Let N be the total number of nodes in the network. Then we define

$$R_0 \;\coloneqq\; \frac{1}{N} \sum_{j=1}^{N} S_T(j) \;=\; \tau \, T \, \frac{1}{N} \sum_{j=1}^{N} \left(\sum_{c} \; \sum_{i \in N^c(j)} f_{ij}^c \right) \;=\; \tau \, T \, \bar{f} \,,$$

where $\bar{f} \;=\; \frac{1}{N} \sum_c \sum_{i,j=1}^{N} f_{ij}^c$ is the average weighted degree of the network. Note that this definition allows us to specify the transmission rate τ in terms of the basic reproduction number R_0, viz. $\tau = R_0/(T\bar{f})$.

To model disease transmission on our extended network, we define the probability for a susceptible individual to become infected on a time interval dt. We assume that $0 < dt \ll 1$ so that infection on our network is unique. Suppose that S_i is susceptible at time t. Then S_i can be infected by any of its neighbors that are infected at time

t in each interaction context c. Thus

$$
\begin{aligned}
&P(S_i \text{ is infected in } [t, t+dt)) \\
&\quad = \sum_{\text{contexts } c} \sum_{\substack{j \in N^c(i) \\ j \text{ infected at time } t}} P(I_j \text{ infects } S_i \text{ through context } c \text{ in } [t, t+dt)) \\
&\quad = \sum_{\text{contexts } c} \sum_{\substack{j \in N^c(i) \\ j \text{ infected at time } t}} \tau \, f_{ij}^c \, dt \\
&\quad = \frac{R_0}{T \bar{f}} \sum_{\text{contexts } c} \sum_{\substack{j \in N^c(i) \\ j \text{ infected at time } t}} f_{ij}^c \, dt \,, \qquad (4.0.1)
\end{aligned}
$$

where we substitute $\tau = R_0/(T\bar{f})$ in the final equality.

As we mentioned in Chapter 2, the frequency f_{ij}^c is a weight assigned to each edge in the sub-network for context c. This approach allows us to account for the different interaction frequencies individuals and their neighbors may have in each context. We then simulate stochastic disease spread on our network, and the state (S, I, or R) of each node is updated at every time step $\Delta t = 1$ day. In the following sections, we refer to the fraction of infected individuals as a function of time as the **epidemic size**, defined as epidemic size $= \frac{I(t)}{N}$.

Since influenza offers long-term immunity and can be distributed through casual interactions (1), we test the spread of a flu epidemic using the discrete SIR model on our extended network. We consider a mean infection time of $T = 4.5$ days as in (28, 29) and use $R_0 = 1.2515$, corresponding to the average of five seasons of flu surveillance data (30). This value for the basic reproduction number is also consistent with estimates in (12, 29). We initialize 0.0034% (10 nodes) of the 3000 individuals in our extended network as infected to provide the seed for disease spreading, while the remaining individuals start as susceptible. The 10 individuals initially infected

consist of a randomly chosen node and its neighbors. If we do not reach our target of 10 infected individuals using this method, we select the remaining nodes by sampling randomly from the individuals connected to the neighbors of the originally infected seed.

4.1 The Importance of a Context-Specific Network Topology

The discrete SIR approach allows for direct comparison with the classical continuous SIR model (27), namely:

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI, \\ \frac{dI}{dt} &= \beta SI - \gamma I, \\ \frac{dR}{dt} &= \gamma I, \end{aligned} \tag{4.1.1}$$

where we specify the same parameters and initial conditions as for the discrete model. We assume a total population of constant size N, where $S(t)$, $I(t)$ and $R(t)$ have the same meaning as in the discrete model and correspond to the sizes of the susceptible, infected, and recovered populations, respectively, at time t. Here, $\beta := \frac{R_0}{NT}$ is the number of new disease cases per unit time and $\gamma := \frac{1}{T}$ represents the rate at which infected individuals recover from the disease (31).

We compare our model results with simulations of equation (4.1.1) for 200 days in Figure 4.1. This timescale allows the epidemic to peak, as well as fully return to its equilibrium (a population with no infected individuals). The differential equation

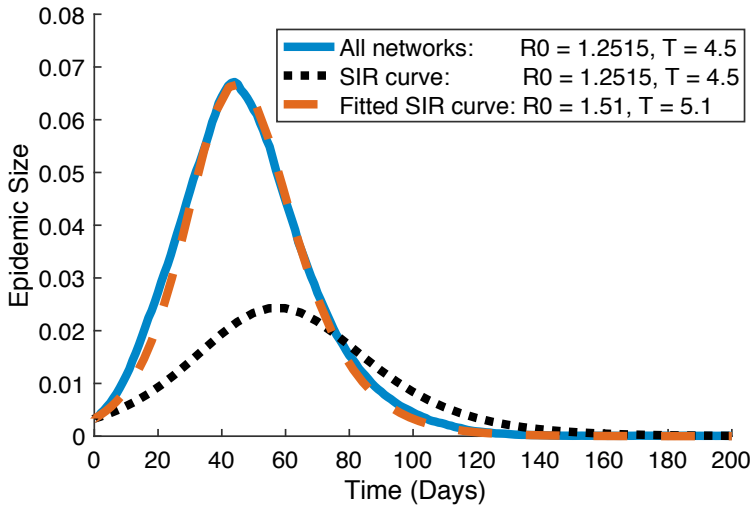

Figure 4.1: Epidemic size over time as predicted by the discrete SIR model (4.0.1) on our full extended network (solid curve), compared to results of the SIR model (4.1.1) with the same R_0 and T parameters (dotted curve) and with fitted parameters (dashed curve). Each of our curves represent the mean over 100 simulations.

model (4.1.1) for influenza leads to a considerably smaller-peak epidemic size (size of the infected population) and a later onset of the disease compared to the discrete SIR model on networks. This means that the model (4.1.1) cannot account for the effects of complex home, social and work interactions on the progression of the disease. On the other hand, our discrete SIR model approach allows us to test the impact of different network structures and interaction contexts on disease spreading. It should be noted that the basic reproduction number R_0 and the mean infection time T in the continuous model (4.1.1) can be chosen and fit so that the epidemic size over time closely resembles our discrete SIR model prediction (Figure 4.1). However, these parameters are different from those considered in our influenza simulations, suggesting that the classical SIR model may yield erroneous parameter estimates for R_0 and T when fitting realistic epidemic data.

4.2 Interaction Context Has a High Impact on Disease Progression

Since the frequencies of interaction are key in the transmission probability formula (4.0.1), we investigate the contribution of different contexts to disease spread in Figure 4.2a. As described in Chapter 3, we represent each interaction context in our network by an individual sub-network with appropriate and unique features. In particular, our home interaction sub-network is highly clustered and disconnected. Our social sub-network, in comparison, is less clustered and more connected. Lastly, the degree distribution for our work sub-network displays a long right tail. By removing the edges in one or more of these sub-networks, we can study how their different features impact the spread of disease.

The horizontal lines in Figure 4.2a correspond to the expected percent contribution of each interaction context to disease spread in the network. We calculate these percent contributions as $\bar{f}_c/\bar{f}$, where c stands for the context (home, work/school, and social) and $\bar{f}_c$ is the average weighted degree of interactions in context c. This measure therefore depends solely on the network structure and frequency of interactions. We obtain the scatter plots in Figure 4.2a by simulating the discrete SIR model for influenza on our network and calculating the percent contribution that infected individuals in each context make to the probability of infecting susceptible nodes at each time step. Figure 4.2a shows that the network predictions and discrete SIR model contributions to infection agree fairly well throughout much of the epidemic lifespan. As expected, the comparison is no longer useful for analysis during the second half of the 200 simulated days when the epidemics dies off (see Figure 4.1).

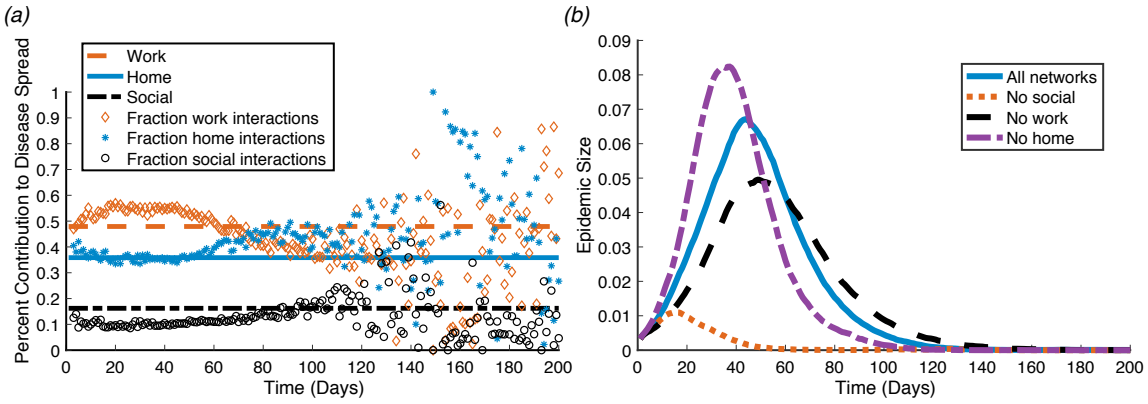

Figure 4.2: Role of interaction context in influenza spread. (a) Percent contribution of each interaction context to disease spread: horizontal lines are expected contributions given the network structure and frequency of interaction, and scatter plots are the contributions to epidemic spread observed by simulating the discrete SIR model (4.0.1) on our network. (b) Epidemic size over time predicted by the discrete SIR model (4.0.1) using our full network (solid curve), compared with removing social interactions (dotted curve), removing work interactions (dashed curve), and removing home interactions (dash-dotted curve). Each of our curves and scatter plots represent the mean over 100 simulations.

Interactions at work have the highest contribution to disease spread until the final days: Figure 4.2a reveals that work sub-network interactions influence disease spread the most as the epidemic grows, peaks, and then begins to die down. This is expected since the interactions in this sub-network are more frequent and are likely to last longer than those in the social context. This result is also consistent with our method of extending the work sub-network in §3.3, which renders the work units more clustered than the social ones. However, in the late days of the epidemic (between days 70 and 80), infections at work become less dominant and home interactions become most responsible for the spread of disease.

Excluding social interactions has the highest impact on disease transmission: We also simulate the spread of influenza on networks where we exclude certain interaction contexts. Figure 4.2b shows how the size of the infected population and the onset of disease are affected when we eliminate the edges in the social, work/school, or home

sub-networks, respectively. Removing connections through the social sub-network impacts disease dynamics most strongly, as it prevents the spread of the epidemic and considerably shortens its duration. This is not surprising given that social interactions are the only connections between more clustered, disparate home and work units. However, this result is not clearly reflected in Figure 4.2a, where the social context has the lowest percent contribution among the networks considered. While there are fewer social interactions compared to work and home encounters, the social context allows the disease to spread across loosely connected clusters in the network, thus facilitating the epidemic.

On the other hand, removing connections through the home sub-network accelerates the outbreak and results in a higher-peak epidemic size. Recall that, on average, interactions at home are the most frequent. Thus when we remove these connections from the network, the average weighted degree of the network $\bar{f}$ decreases considerably in equation (4.0.1). Since R_0 remains constant in all of our simulations, this produces an increased transmission risk τ for each susceptible-infected contact. Nevertheless, this result is expected given the structure of our extended home sub-network, which consists of disconnected, fully clustered home units. Without these isolated interactions, individuals are likely to increase their social and work interactions, which would allow the infection to spread more freely across the connections in the network. Thus, by increasing the clustering in the network, the home sub-network limits the overall disease transmission.

4.3 Dynamic Responses to Disease Substantially Reduce Epidemic Size

The network-based discrete SIR model allows us to test how dynamic responses in the population alter the duration and onset of disease. We consider a few realistic reactions to the onset of influenza, such as a scenario in which home interactions become more frequent following infection, while social and work interactions are reduced. We model such responses to disease by lowering the interaction frequencies of individuals 1–2 days after they become infected.

Considerably reducing interactions at work leads to a smaller epidemic size and duration of infection: We show the effect of reducing interactions through different sub-networks in Figure 4.3a. Reducing social interactions to a large extent decreases the size of the epidemic, but predicts a similar or slightly increased epidemic duration (dash-dotted and starred curves, Figure 4.3a). On the other hand, greatly decreasing the frequency of interactions in the work context yields a smaller epidemic size and duration of infection (dashed and dotted curves, Figure 4.3a). This is expected given the network structure that we considered. In particular, reducing frequent work interactions does not allow the spread of the infection inside work clusters; and this, in turn, limits the spread of the disease to other network clusters through occasional social encounters.

This observation is also supported by Figure 4.3b, where we plot the percent contributions to infection through each context in the situation where work interactions are removed completely after infection onset. Compared to Figure 4.2a, the work and home percent contributions switch, with home clusters becoming the most influential in disease spread. The social contribution increases to a small extent, but

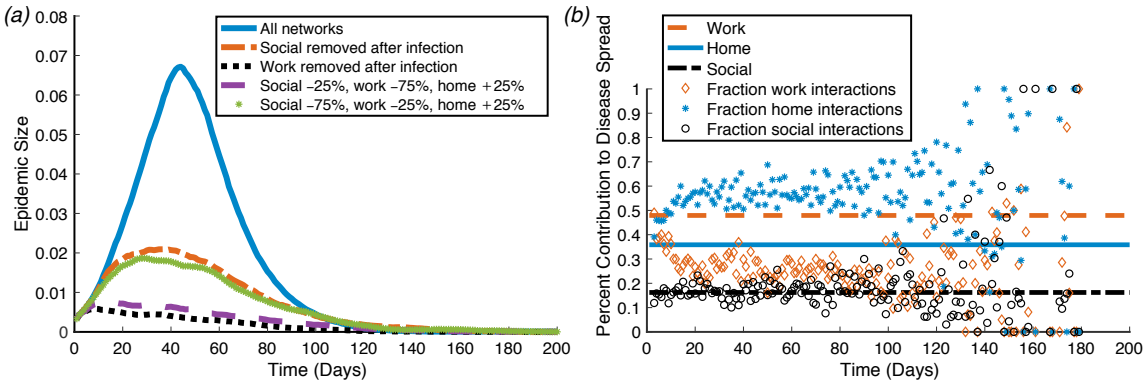

Figure 4.3: Impact of dynamic context-specific responses on disease spread. (a) Epidemic size over time predicted by simulating the discrete SIR model (4.0.1) on our full network in comparison to simulations incorporating various dynamic network responses 1–2 days after disease onset. Percentages represent proportional changes in interaction made after infection (e.g., "Social: −25%" means an infected individual will remove 25% of their usual social interactions). (b) Percent contributions of each context to disease spread: horizontal lines are as in Figure 4.2a, and scatter plots are contributions to epidemic spread observed by simulating the discrete SIR model (4.0.1) under the scenario that individuals remove work interactions after infection. Each of our curves and scatter plots represent the mean over 100 simulations.

only a few of these interactions are likely to spread the disease, as individuals do not interact in work clusters and can thus no longer spread the infection through social interactions as well. The high variability of context contributions in this figure is due to the small epidemic size in this dynamic reaction to influenza (dotted curve, Figure 4.3a).

4.4 Model Results Capture Trends in NHS Flu Call Data

Our extended network was constructed using the interaction structure from a diary-based study in a university setting (1) and likely reflects characteristics of a small population of 3000 people. Accordingly, we do not expect our model to fit flu dy-

namics data for the large population of England. However, in order to get a rough reference of the epidemic sizes and peaks typically observed, we compare the epidemic size curves predicted by our discrete SIR model with data from the Public Health England (PHE) real-time syndromic surveillance system (2), which provides weekly reports from October to May. Specifically, we extract information on the percent of National Health Service (NHS) 111 calls attributed to cold or flu during the winter (2) using WebPlotDigitizer (32).

Figure 4.4 overlays information on the fraction of NHS 111 calls (2) for cold and flu in England with simulation dynamics given different responses to the epidemic. Since the seasonal data in (2) varies in shape every year, we plot a few recent representative examples of this NHS data. It is worth noting that we shifted all the curves on the time axis so that the epidemic size peaks are aligned across simulation and NHS data. In Figure 4.4b, we also shift our discrete SIR model results on the epidemic size axis by a constant to account for the fact that realistic outbreak data for influenza has a background epidemic size even outside the peak epidemic weeks. These shifts do not affect our comparisons, since we are primarily interested in the peak epidemic size and the epidemic duration.

Reducing social and work interactions yields epidemic dynamics on similar scales with flu season data: Figure 4.4a shows that simulating influenza spread across our full network without including any dynamic responses to disease is likely overestimating the epidemic size. We also note that in our simulated model, the infection outcomes include both symptomatic and asymptomatic cases, whereas the NHS data only reflects symptomatic cases. The peak epidemic size predicted by our model in this setting is larger than the proportions of NHS 111 calls for cold and flu recorded for all of the flu seasons in 2013–2019 (2). On the other hand, incorporating dynamic changes in the social sub-network after infection predicts epidemic behavior that

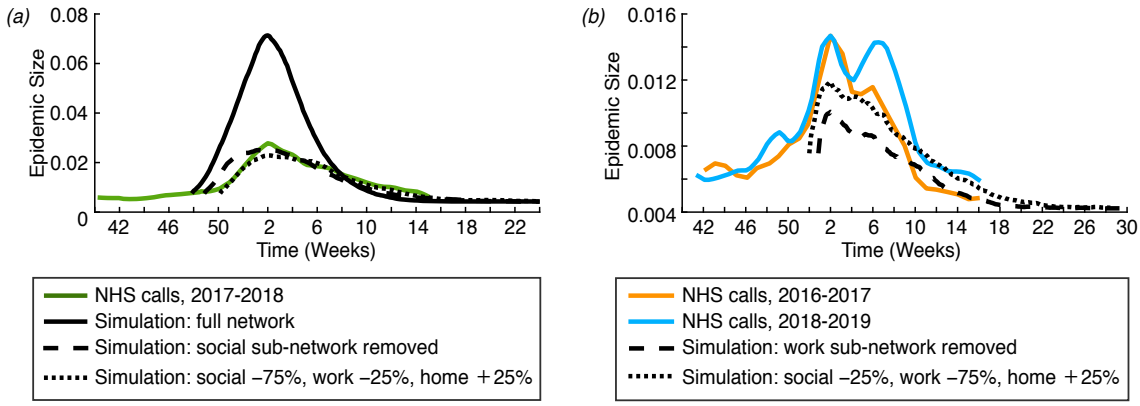

Figure 4.4: Comparison of our results with influenza data (2) from England. (a) Epidemic size evolution predicted by the discrete SIR model (4.0.1) using various dynamic responses 1–2 days after disease onset (dashed and dotted black curves) and flu call data (green curve) reported by PHE based on 2017–2018 NHS 111 calls (2). For comparison, we also show how the SIR model behaves on our full network in the solid black curve. (b) Epidemic size evolution predicted by the discrete SIR model (4.0.1) using various dynamic responses 1–2 days after disease onset (dashed and dotted black curves) and example flu call data (orange and blue curves) reported by PHE based on 2016–2017 and 2018–2019 NHS 111 calls (2). Percentages in (a) and (b) represent proportional changes in interaction made after infection, (e.g., "Social: -25%" means an infected individual will remove 25% of their usual social interactions). Each of our curves represent the mean over 100 simulations.

closely resembles the 2017–2018 flu season, which was characterized by moderate to high levels of influenza activity (2). Similarly, large changes in the pattern of work interactions leads to a good prediction of the start of the epidemics during the 2016–2017 and 2018–2019 seasons, both characterized by low to moderate flu activity (see Figure 4.4b). Our model, therefore, suggests that strategies involving reductions in work interactions after flu onset may be most effective at avoiding severe influenza seasons. Our results also indicate that, as expected, individuals are likely to change their social and work interactions shortly after they contract the flu.

We note that we do not expect to recreate the double-peak curves observed in some of the seasons represented in Figure 4.4, since the dynamic responses in our model are not influenced by factors such as cognition of epidemic spread (as in (12)). In the study (12), the authors use a heterogeneous graph-modeling approach to de-

scribe flu virus transmission in a population of hospital patients and an agent-based model incorporating many unknown parameters to model the dynamic change in individuals' interactions as a reaction to the epidemic. Although our minimal SIR modeling approach does not reproduce all of the features of realistic epidemic dynamics in Figure 4.4 (such as double peaks or onset of the epidemics), our simulations are similar to real influenza epidemics in terms of outbreak size and duration. Moreover, our focus in this work is on developing methods for lifting interaction-based diary data to larger networks. In the future, it would be interesting to explore how more realistic disease models (e.g., that include dynamic responses based on cognition) behave on our lifted network in comparison to influenza data.

Chapter Five

A Context-Specific Network for Providence, Rhode Island

In Chapter 3, we presented an algorithm for generating an artificial network which extends the interaction structure recorded in (1). Our algorithm represented each interaction context with a separate sub-network in order to preserve many of the features that distinguish the structure in home, social, and work settings. For example, the home interaction sub-network is highly clustered and disconnected, while the social sub-network, in comparison, is less clustered and more connected. In Chapter 4, we simulated an influenza outbreak on a hypothetical network and discovered the importance of taking a context-specific approach to network construction (see Figure 4.1). In fact, Figure 4.4 illustrated that the simulations on this context-specific network were able to capture trends observed in NHS flu call data.

In this chapter, we adapt the algorithm introduced in Chapter 3 to generate a specific, real-world network which accounts for the context-specific interaction features that are recorded in the Warwick data (1). In particular, we build a larger network of 100,000 nodes to model the city of Providence, Rhode Island (RI), which has an estimated population of 179,883 individuals (33). Recall from Chapter 3 that each sub-network was built in two main steps. First, edges were assigned between nodes to form a collection of units. Second, we assign a weight to each edge reflecting the frequency of interaction in that context, which we sample from the appropriate frequency distribution for the Warwick data (1). We now alter step one: in order to build a network suitable for Providence, RI, we form units which match the distribution of homes and businesses in the United States (U.S.) and the state of Rhode Island, respectively. Then, step two remains the same, that is, we allow the interaction structure of the diary data (1) to inform the frequency of interaction within each unit. Once again, we do not differentiate between the proximity of interaction (casual or skin-to-skin) and we will therefore focus on the spread of diseases which do not require skin-to-skin contact for transmission in the remaining

chapters.

In the sections below, we outline the modifications made to the algorithm in Chapter 3 to generate a network for Providence, RI and present the steps for constructing each sub-network. The home, social, and work/school sub-networks are discussed in §5.1, §5.2, and §5.3, respectively.

5.1 Home Network Construction

In Section 3.1, we built the home network by adding one home unit at a time. First, we sampled a household size $n+1$ and generated a clique containing $n+1$ nodes. In order to build a network that is suitable for Providence, RI, we sample each new home unit size from the size distribution of households in the U.S. in 2019, which we obtain from U.S. census data (34). Then, we assign an interaction weight to each edge in the unit as we did before: by sampling from the frequency distribution for home interactions in the Warwick data (1). We add this new home unit to the existing network and begin the process again. The resulting home sub-network consists of disconnected, fully clustered households which match the size distribution of homes in the U.S. and interaction frequencies on each edge that were drawn from the diary-data recorded in (1).

We summarize these steps below:

1. Sample a home unit size $n+1$ from the household size distribution in the U.S. (34).
2. Generate a clique of size $n+1$.

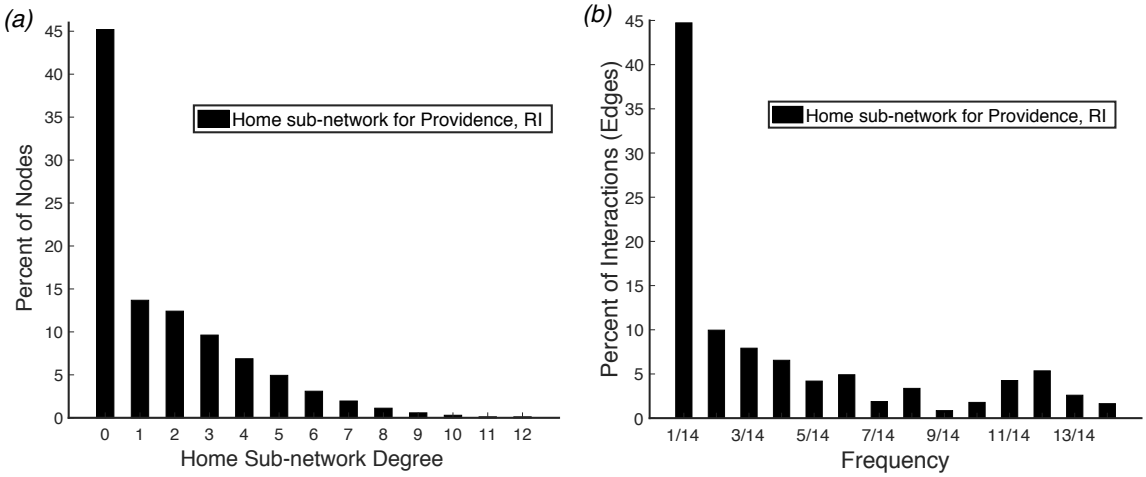

Figure 5.1: (a) Degree and (b) frequency distributions for the home sub-network for Providence, RI. We use U.S. household-size data (34) and the frequency distribution for home interactions in the Warwick data (1) to construct the network.

3. Assign an interaction weight to each edge in the home unit by sampling from the appropriate home frequency distribution of the Warwick data (1): the distribution for $n \leq 8$ or for $n > 8$.

4. Add the new home unit to the existing network and repeat from Step 1 until we have generated $100,000$ nodes (adjusting the size of the final home unit as necessary).

We plot the degree and frequency distributions for the home sub-network in Figure 5.1.

5.2 Social Network Construction

In Section 3.2, we constructed the social network in three steps, taking into account the key features of the Warwick data (1) and measurements (20, 21) of clustering in online social communities. In particular, we built identical social units of uniform

degree to match the approximately uniform character of the Warwick social degree distribution (1). Then, we added and shuffled edges in the network to account for high-degree individuals and to match the clustering coefficient observed in online communities (21). We assume that these features exhibited by real-world social interactions are not inherently location-specific, and thus they can be used to model social interactions in many locations. Therefore, we construct the social network for Providence, RI using the same steps listed in Section 3.2. However, since our social network contains $100,000$ nodes, we must shuffle far more edges in the network to reach the desired average local clustering coefficient $C = 0.16$ reported by Ahn *et al.* (21). Specifically, we found that $C = 0.16$ when we shuffle $M = 285,000$ edges.

These updated steps are outlined as follows:

1. Generate 2631 identical social building blocks of size 38 so that the degree distribution within each social unit is approximately uniform from 0 to 36. (Degree 18 necessarily appears twice.)
2. For each social unit, choose $\alpha = 1$ of the $\beta = 5$ most popular (highest degree) nodes in the unit. Connect each of the chosen nodes to $\gamma = 25$ high-degree nodes randomly selected from other social units, where "high-degree" means one of the top β highest degree nodes in a social unit.
3. Randomly select $M = 285,000$ edges and, for each such edge, disconnect one end and reconnect it to another node chosen at random.
4. Randomly select and remove 16 nodes to reduce the total network size to the target $100,000$ nodes.
5. Assign interaction weights to each edge by sampling from the appropriate social frequency distribution for the Warwick data (1): the distribution for $n \leq 18$ or for $n > 18$.

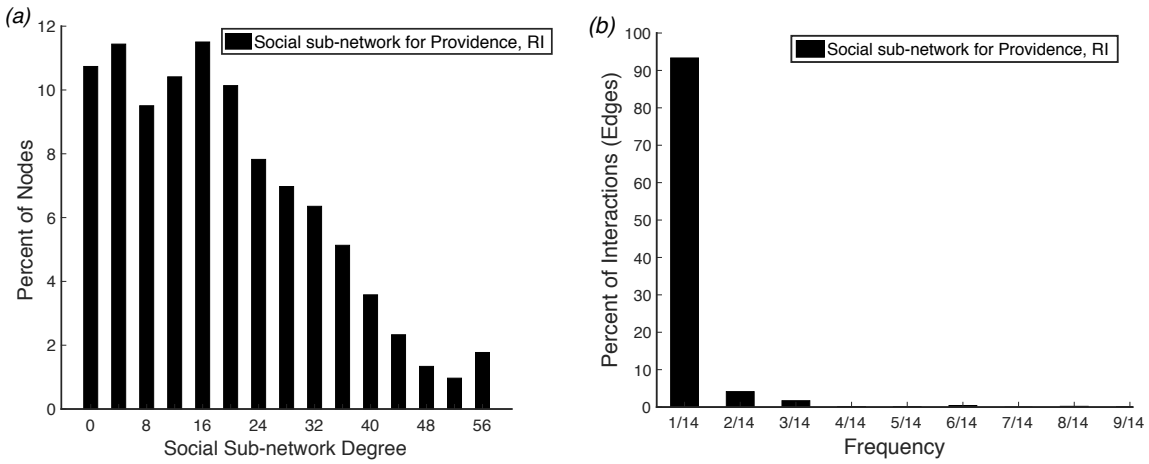

Figure 5.2: (a) Degree and (b) frequency distributions for the social sub-network for Providence, RI. We use measurements of clustering in online social communities (20, 21) and the frequency distribution for social interactions in the Warwick data (1) to construct the network.

The degree and frequency distributions for the social sub-network are shown in Figure 5.2.

5.3 Work Network Construction

In Section 3.3, we built the work network by first forming a collection of unit sizes according to the size distribution of businesses in Coventry, UK. Since our goal in this section is to build a work network suitable for Providence, RI, we use U.S. census data on business sizes in the state of Rhode Island in 2017 (35) to construct unit sizes in a $100,000$-person network. Then, we assign the edges in each unit the same way we did in Section 3.3. Specifically, we generate the large work units using the original Barabási-Albert algorithm (24) and we generate the small work units such that the degree distribution within each unit is roughly uniform. Finally, we assign a frequency of interaction to each edge in the network by sampling from the work frequency distribution for the Warwick data (1).

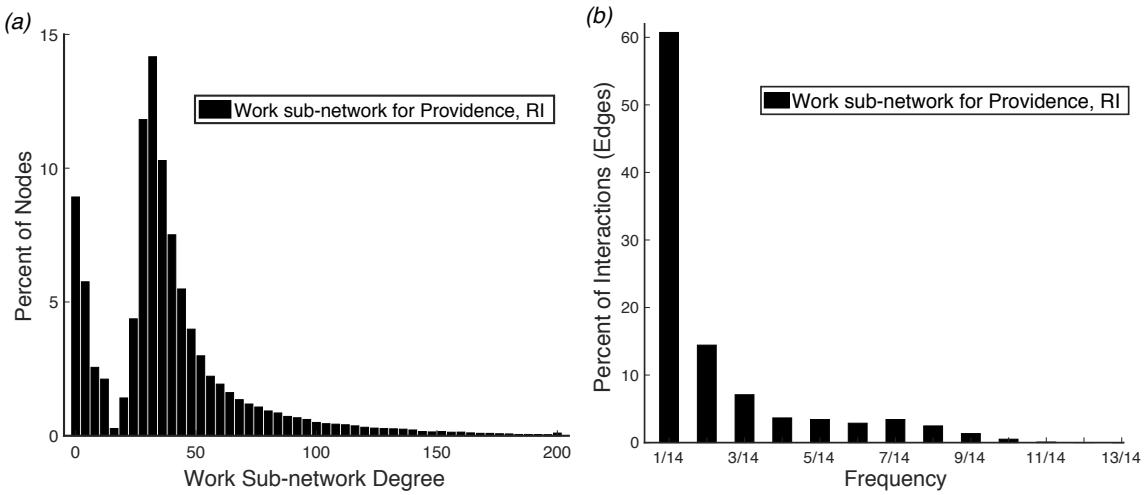

Figure 5.3: (a) Degree and (b) frequency distributions for the work sub-network for Providence, RI. We use Rhode Island business-size data (35) and the frequency distribution for work interactions in the Warwick data (1) to construct the network.

We summarize these new steps below:

1. Use the original Barabási-Albert model (24) with $m = m_0 = 30$ to generate the following large work units:

   - 79 work units made up of 600 nodes
   - 50 work units made up of 300 nodes
   - 312 work units made up of 60 nodes

   (We choose this value of m to best match the degree distribution of the real data (1).)

2. Generate the following small work units with a uniform degree distribution in each:

   - 480 work units made up of 16 nodes, each with a degree distribution that is roughly uniform from 0 to 14 (degree 7 will appear twice)
   - 750 work units made up of 8 nodes, each with a degree distribution that is roughly uniform from 0 to 6 (degree 3 will appear twice)

- 1300 work units made up of 4 nodes, each with a degree distribution that is roughly uniform from 0 to 2 (degree 1 will appear twice)

3. Together the work/school units generated in Steps 1–2 represent the network. Assign an interaction weight to each edge in the network by sampling from the work frequency distribution for the Warwick data (1).

We present the degree and frequency distributions for the work sub-network in Figure 5.3.

Chapter Six

COVID-19 Dynamics in Providence, Rhode Island

We return to the discrete susceptible-infected-recovered (SIR) model laid out in Chapter 4 to explore the spread of an infectious disease on our context-specific network for Providence, RI. We use the same notation as in Chapter 4, denoting the i^{th} susceptible node by S_i and the j^{th} infected node by I_j. As before, we assume each infected node recovers from the disease after a time drawn from an exponential distribution with mean T, where T is the average duration of infectiousness. We define the basic reproduction number to be $R_0 = \tau\, T\, \bar{f}$, where τ is the transmission risk defined in Chapter 4 and $\bar{f}$ is the average weighted degree of the network. In words, R_0 represents the average number of secondary infections produced from one infected individual during the infectious period T in a population that is otherwise fully susceptible.

We define the probability for a susceptible individual to become infected on a time interval $0 < dt \ll 1$. Let $N^c(i)$ be the set of indices j such that the jth node is connected to node i in interaction context c. Then

$$P(S_i \text{ is infected in } [t, t+dt)) = \frac{R_0}{T\bar{f}} \sum_{\text{contexts } c} \sum_{\substack{j \in N^c(i) \\ j \text{ infected at time } t}} f_{ij}^c \, dt\,, \tag{6.0.1}$$

where f_{ij}^c is the interaction frequency between S_i and I_j in context c. This derivation is treated in more detail in Chapter 4. We then simulate stochastic disease spread on our network for Providence, RI, updating the state (S, I, or R) of each node at every time step $\Delta t = 1$ day. In the following analyses, we refer to the cumulative total infections over time as the **total infection count**, defined as total infection count $= I(t) + R(t)$. This quantity represents the total number of infections caused by the outbreak by time t and does not remove "closed cases" (that is, those who have recovered or died).

Since coronavirus disease 2019 (COVID-19), the disease caused by severe acute respiratory coronavirus 2 (SARS-CoV-2), is distributed through casual interactions (36), we use this discrete SIR model on our network for Providence, RI to study a COVID-19 outbreak in Providence, RI. Since the COVID-19 pandemic is still developing, we restrict our study to the start of an outbreak in this chapter and then explore the effect of control measures which are introduced early in the development of the epidemic in Chapter 7. We also do not differentiate between asymptomatic and symptomatic individuals since the infectiousness in each case is similar (37).

At this time, the duration of conferred immunity to SARS-CoV-2 after recovering from the infection is still unclear. However, a very recent study tested the immunological memory to SARS-CoV-2 in recovered individuals and found that 95% of the test subjects retained immune memory six months after infection (38). Thus we assume that immunity will last at least 200 days in all recovered individuals and utilize the SIR model framework (6.0.1) on this timescale (27). In the future, as more is understood about the nature of this disease and the duration of conferred immunity, our model can be updated or exchanged. For example, if the virus is found to confer immunity for a fixed and finite period of time, we could model a COVID-19 outbreak on our network for Providence, RI using an SIRS model, which accounts for short-term immunity and then a return to susceptibility.

We begin our simulation with a single infected node, chosen at random, while all other individuals in our network for Providence, RI are initialized as susceptible. Several studies suggest that individuals with mild to moderate COVID-19 remain infectious up to 10 days after symptom onset and individuals with severe to critical COVID-19, or acute immunocompromise, remain infectious up to 20 days after symptom onset (39, 40, 41, 42, 43). We also note that the infectious period for COVID-19 begins at least 2 days before the onset of symptoms (44). With these

details in mind, we consider a mean infection time of $T = 14$ days. Finally, to determine the appropriate value of R_0 for our purposes, we calibrate our model using infection data from Providence, RI during the COVID-19 pandemic in Section 6.1 below.

6.1 Calibrating the Basic Reproduction Number R_0

In the discrete SIR model (6.0.1) described in the section above, we define the basic reproduction number R_0 to be $R_0 = \tau T \bar{f}$, where τ is the transmission risk defined in Chapter 4 and $\bar{f}$ is the average weighted degree of the network. In words, R_0 represents the average number of secondary infections produced from one infected individual during the infectious period T in a population that is otherwise fully susceptible. Thus we can think of R_0 as a measure of the contagiousness, or transmissibility, of a pathogen. The value of R_0 is not constant for a given pathogen; instead it is highly sensitive to human behavior and the region for which it is calculated (45). Currently, the Centers for Disease Control and Prevention (CDC) reports that $R_0 = 2.5$ is the best estimate for the SARS-CoV-2 reproduction number in the U.S. (46). This estimate is representative of the point estimates of R_0 obtained in (47, 48, 49, 50, 51), but the value is expected to change as more data becomes available. In general, however, the estimation of R_0 tends to be model-specific. That is, R_0 is determined relative to the modeling framework being used at the time and the value may even differ between models for the same disease (52).

Our goal is to determine an appropriate value for R_0 in our SIR model (6.0.1) in order to approximate the transmission rate of SARS-CoV-2 in Providence, RI. The

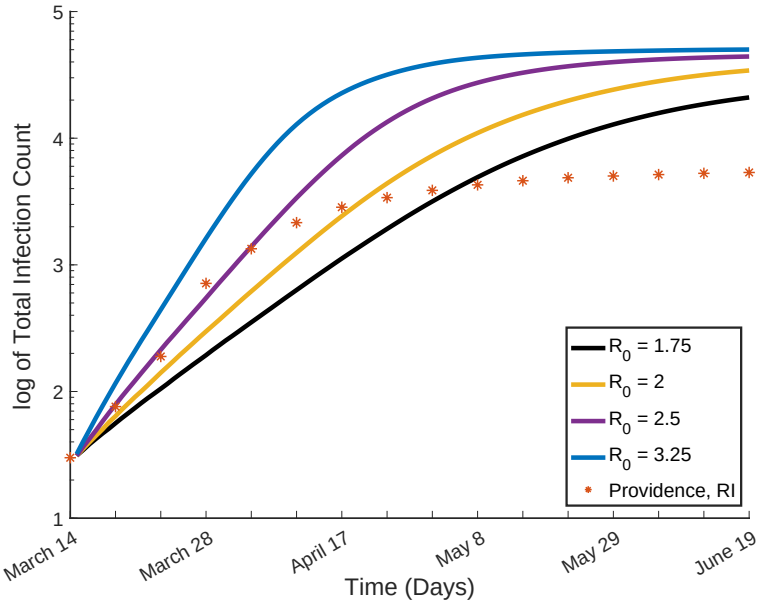

Figure 6.1: Total infection count over time predicted by simulating the discrete SIR model (6.0.1) on our network for Providence, RI given various values of R_0. The scatter plot represents the weekly total infection count recorded in Providence, RI by the RIDOH (58). We remove the initial data point when Providence, RI recorded < 5 infections and plot from March 14 (when Providence, RI recorded 30 infections). We align this date with the day our simulations surpass 30 total infections. Each of our curves represent the mean over 100 simulations.

value of R_0 should represent the basic reproduction number during baseline spread, that is, without any control measures present in the city. To do this, we run the model repeatedly, varying the value of R_0 in each simulation and observing the total infection count of the resulting COVID-19 outbreak on our network for Providence, RI. Considering several estimated ranges for R_0 in the literature (53, 54, 55, 56, 57), we begin with $R_0 = 1.75$ and increase the value of R_0 by 0.05 in each simulation until we reach $R_0 = 3.5$. Then, we plot the total infection count over time for each simulation and compare the results to the actual recorded infections in Providence, RI during the COVID-19 pandemic. For the latter, we use weekly infection data obtained from the Rhode Island Department of Health (RIDOH) for the city of Providence, RI between March 1, 2020 and September 11, 2020 (58).

There are a few limitations to this method. First, the infection data (58) recorded in Providence, RI during the COVID-19 pandemic is actually a measure of positive test results and may not reflect the true number of cases in Providence, RI at the time. Thus, our model will be calibrated to the number of positive test results, rather than the actual number of infections in the city. Given more accurate data for the total infections in Providence, RI during the COVID-19 pandemic, we could re-calibrate our model using the same method as above in order to obtain a better estimate for R_0. Second, infection control measures were implemented early in the COVID-19 pandemic in Providence, RI (see Section 7.1). In fact, the first infection was recorded in Rhode Island on March 1, 2020 (58) and then widespread stay-at-home orders and business closures were introduced on March 28, 2020, when the city had only surpassed 100 total infections (viz. 0.06% of the population) (59). Therefore, the infection data (58) that we use to calibrate our model only logs uninterrupted disease spread for 28 days, which does not allow us to match the results of our test simulations with the data deep into the epidemic lifespan.

Figure 6.1 presents the total infection count predicted by simulating the SIR model (6.0.1) with several different values of R_0. We find that when $R_0 = 2.5$, our model (6.0.1) approximates the infection count in Providence, RI (58) at the start of the outbreak fairly well. Interestingly, $R_0 = 2.5$ is also the current best estimate for the SARS-CoV-2 basic reproduction number in the U.S. according to the CDC (46). For these reasons, we assign $R_0 = 2.5$ in our model (6.0.1) for all subsequent simulations.

6.2 Work Interactions Maintain Disease Transmission

We examine the contribution of each interaction context in our network for Providence, RI to the spread of COVID-19 in Figure 6.2a. The expected percent contributions, based solely on the network structure, predict the actual contributions realized by simulating the discrete SIR model (6.0.1) on our network for COVID-19 quite well during the first 200 days of the outbreak. For reference, the calculations of these measures are described in Section 4.2. Figure 6.2a reveals that work interactions contribute most to the spread of disease, which is not surprising since interactions at work occur more frequently and likely last longer than social interactions. This also aligns with the construction of the work sub-network, which renders the work units more clustered than the social units. Interestingly, the home sub-network interactions contribute least to disease spread on the network for Providence, RI. Although interactions at home are more frequent than social and work interactions, they tend to be contained within smaller, disconnected groups. In particular, we construct our home units using household data for the U.S. in 2019 (34) where 90% of homes contained 4 or fewer individuals (with single-occupant homes comprising one-quarter of those).

To better understand how the unique features of our sub-networks influence disease progression, we simulate the spread of COVID-19 on networks where we exclude certain interaction contexts. Figure 6.2b shows how the total number of infections and the rate of new infections in the population are affected when we eliminate the edges in the home, social, or work/school sub-networks, respectively. Removing connections in the home sub-network slows the rate of infection during the course of the epidemic, but it does not strongly impact the total number of individuals

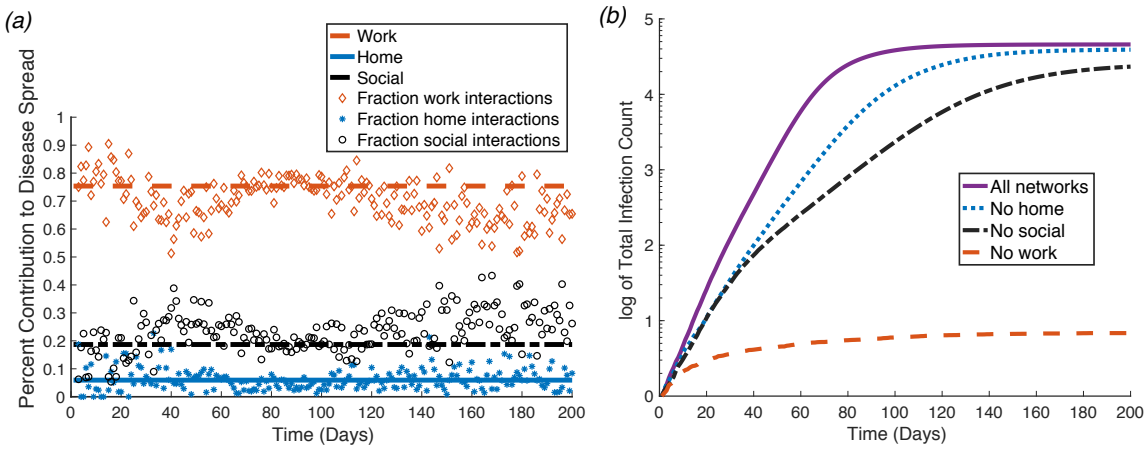

Figure 6.2: Role of interaction context in COVID-19 dynamics. (a) Percent contribution of each interaction context to disease spread: horizontal lines are expected contributions given the network structure and frequency of interaction, and scatter plots are the contributions to epidemic spread observed by simulating the discrete SIR model (6.0.1) on our network. (b) Total infection count over time predicted by the discrete SIR model (6.0.1) using our full network (solid curve), compared with removing home interactions (dotted curve), removing social interactions (dash-dotted curve), and removing work interactions (dashed curve). Each of our curves and scatter plots represent the mean over 100 simulations.

who will become infected. This is expected since most infections occur at work or through social interactions during the first 200 days of the outbreak (see Figure 6.2a). Reducing social interactions slows the spread of disease more substantially, however, since the social interactions act as bridges between more loosely connected clusters in the network. Yet removing interactions at work impacts disease dynamics most strongly, as it prevents the spread of the epidemic and considerably shortens its duration. Nevertheless, we find that removing interactions in each context slows the spread of disease on the network. This is noteworthy for scenarios where it is important to minimize the number of individuals infected at a given time.

Chapter Seven

Intervention Strategies and the Spread of COVID-19

We saw in Chapter 4 that the network-based discrete SIR model allows us to explore how dynamic responses in the population alter disease progression. Here we study the impact of several infection control measures implemented during a COVID-19 outbreak in Providence, RI. Due to the interaction-based structure of our network for Providence, RI, we focus on control measures which reduce interactions between individuals to curtail the spread of disease.

In Section 7.1, we examine some of the measures taken in Providence, RI during the COVID-19 pandemic. We look specifically at the stay-at-home order, which limited interactions in social settings, and the closure of all non-essential businesses, which limited interactions in work settings (59). These measures were implemented simultaneously on the day that the total infection count in Providence, RI surpassed 100 infections (58). Then, in Section 7.2, we consider a partial business closure, in which small businesses remain open while large businesses are required to close. Lastly, we study a contact tracing scheme, where infected individuals and their collection of contacts in the network self-quarantine (Section 7.3). This effort works to isolate known infected individuals, as well as individuals who may have been infected, thereby reducing interaction during the infectious period.

7.1 Lockdown in Providence, RI during the COVID-19 Pandemic

During the COVID-19 pandemic, several measures were implemented in Providence, RI to limit the interactions between individuals. Two such measures began on March 28, 2020 after Providence, RI had recorded more than 100 total infections: a stay-at-home order, which limited interactions in social settings, and a closure of all

non-essential businesses, which limited interactions in work settings (59). These measures remained in place until May 8, 2020 (60). For brevity, we call this six-week period the **lockdown period**.

Similar measures have been discussed in recent studies and the lowered contact rate which results from these widespread control measures has been modeled in various ways. For example, the authors in (61) incorporated a time-dependent transmission rate into their modified SEIR model which decreases after the introduction of the lockdown measures. Other works construct separate models for the different stages of lockdown, including one representing disease transmission before the lockdown is implemented and another representing disease transmission during the lockdown. The transmission rate is then defined to be lower in the lockdown model (62, 63). Another compartmental model separates the susceptible and infected individuals into two classes: those who reduce their interactions due to the lockdown measures and those who do not reduce their interactions. Then, the impact of the intervention is quantified by assigning a separate, lower infection rate to the class of individuals who participate in the lockdown (64).

The above examples (61, 62, 63, 64) and several other studies use the lowered infection rate which results from the intervention as a way to represent the decreased interaction rate in the community. In this section, we center on the decreased interaction rate instead and determine the extent to which we must lower interactions in our network in order to reproduce the infection data recorded in Providence, RI during the lockdown period (58). To do so, we simulate a COVID-19 outbreak on our network for Providence, RI using the SIR framework (6.0.1) laid out in Chapter 6. We begin the simulation without any intervention. Then, in order to match the timeline of the lockdown period, we lower the interaction frequencies in our network on the day that the outbreak surpasses 100 or more total infections. We consider this

Percent reductions on interaction frequencies throughout network			
	Social	Work	Home
Week 1 (Day 1)	−5%	−5%	+10%
Week 2 (Day 8)	−10%	−5%	+10%
Week 3 (Day 15)	−15%	−20%	+20%
Week 4 (Day 22)	−25%	−35%	+20%
Week 5 (Day 29)	−35%	−50%	+5%
Week 6 (Day 36)	−25%	−50%	+5%

Table 7.1: Weekly changes in interaction which reproduce the total infection count in Providence, RI during the lockdown in 2020 (58). Percentages represent proportional changes in interaction made at the start of each week for all individuals (e.g., "Social: −5%" means all individuals will remove 5% of their current social interactions). "Day 1" of the intervention corresponds to the day that the total infection count surpasses 100 infections. Reductions are applied cumulatively. After Week 6, the resulting interaction frequencies are maintained for the remainder of the simulation.

to be "Day 1" of the intervention. We then lower the social and work interactions at the start of each week for six weeks. We also account for a resultant increase in home interactions at each week. Then, after the sixth week, the lowered interaction frequencies are maintained for the remainder of the simulation.

The specific proportion of interactions in our network which must be lowered in order to reproduce the total infection count in Providence, RI during the lockdown period (58) are listed in Table 7.1. Note that the reductions are applied cumulatively. For example, at the start of Week 1, we lower the social interaction frequencies throughout the network by 5%. Then, at the start of Week 2, we lower the social interaction frequencies throughout the network by a further 10%, which amounts to a 14.5% reduction on the original social interaction frequency values.

In Figure 7.1a, we plot the effects of these changes iteratively along with the infection data for Providence, RI during and after the lockdown period (58). Each curve represents a COVID-19 outbreak in which some or all of the weekly reductions are implemented. When we reduce the interactions in the network for the full six-

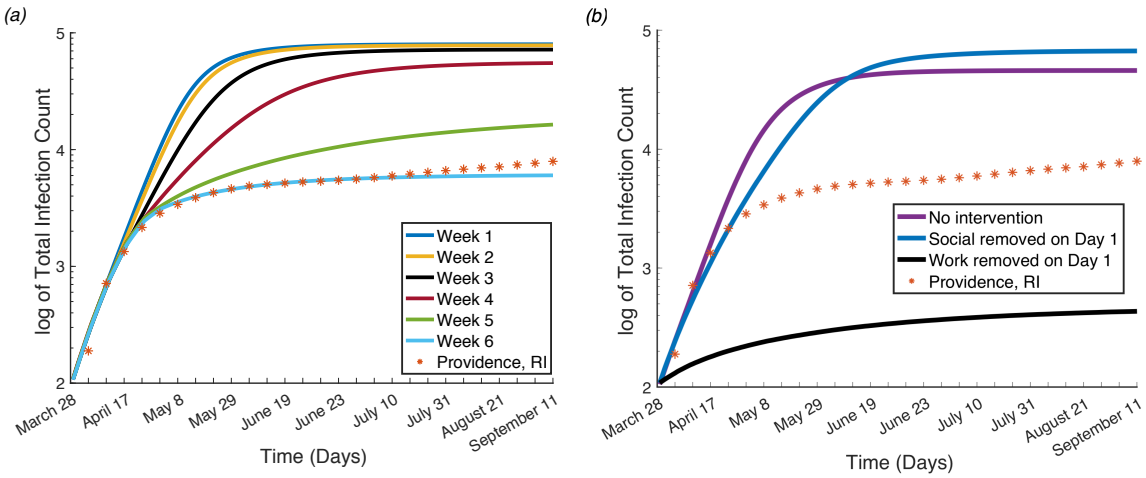

Figure 7.1: Relationship between interaction rates and disease transmission during the lockdown in Providence, RI. (a) Total infection count over time predicted by simulating the discrete SIR model (6.0.1) on our network for Providence, RI, along with scenarios in which all or some of the weekly reductions in Table 7.1 are implemented. "Week x" indicates that the changes in interaction listed in Table 7.1 were implemented from Week 1 through Week x and then maintained for the remainder of the simulation. (b) Total infection count over time predicted by simulating the discrete SIR model (6.0.1) on our network for Providence, RI, along with scenarios in which one interaction context is removed at the start of the intervention. The scatter plot represents the weekly total infection count recorded in Providence, RI by the RIDOH (58). We align the date March 28 on the horizontal axis with Day 1 of the intervention in each of our simulations. Day 1 occurs when the total infection count has surpassed 100 infections. Each of our curves represent the mean over 100 simulations.

week lockdown period (the "Week 6" curve, Figure 7.1a), the predicted total infection count agrees with the infection data from Providence, RI (58) for about 119 days (which includes the six-week lockdown period, as well as the following 7 weeks). To match the data further, we could continue the process, increasing the interactions in the network appropriately.

Interactions were significantly reduced during the lockdown period, with the greatest decrease occurring in work settings: In order to generate a COVID-19 outbreak with a total infection count that matches the data recorded in Providence, RI (58), we made significant reductions to the interaction frequencies within our network in

both social and work settings. Interactions were lowered the most during the later weeks of the six-week lockdown. This difference is reflected in Figure 7.1a, where we plot the total infection count of outbreaks where only some of the weekly reductions are implemented. Indeed, when the social and work interactions were only reduced by the values in Table 7.1 for the first one, two, and three weeks, the resulting total infection counts show little variation. However, larger decreases in the interactions which occurred in the final three weeks of the lockdown swiftly lowered the total number of infections caused during the first 200 days of the outbreak. We also notice that the most significant decrease occurred in work interactions. This could indicate that the closure of all non-essential businesses was more wide-reaching than the stay-at-home order.

A decrease in work interactions is necessary to reproduce the total infection count during the lockdown period in Providence, RI: We also tested a scenario in which all social interactions in the network were removed completely at the start of Week 1 and for the remainder of the simulated outbreak (blue curve, Figure 7.1b). In this scenario, no changes were made to the work interactions. Figure 7.1b shows that this hypothetical response results in an outbreak with a total infection count that is far greater than the total number of infections recorded in Providence, RI during and after the lockdown period (58). Thus, to reproduce the lowered infection rate observed in Providence, RI during this time, some reduction of the work interactions is necessary.

7.2 Small-Business Interactions Only Mildly Impact Disease Progression

In Section 7.1, we discussed some of the infection control measures implemented during the COVID-19 pandemic in Providence, RI to slow the spread of disease. One of these was a closure of all non-essential businesses aimed at reducing the interactions between individuals in work settings (59). By simulating a COVID-19 outbreak on our network for Providence, RI, we saw that a decrease in the frequency of work interactions was necessary to achieve the lowered infection rate (58) recorded in Providence, RI during the lockdown (see Figure 7.1b). However, the economic impact of long-term closures can be substantial on businesses which cannot continue operating through remote or virtual means. Oftentimes larger businesses, such as universities and corporations, are able to operate remotely, while smaller businesses cannot. In an effort to mitigate this economic impact on smaller businesses, we consider the effect of a partial business closure on the evolution of COVID-19. We explore several closure scenarios which emphasize the larger businesses and leave smaller businesses open.

Recall that the work sub-network in our network for Providence, RI is comprised of many disconnected work units, ranging from 4 to 600 employees (see Section 5.3). For our study, we define units with 60 or fewer employees to be "small businesses" and units with greater than 60 employees to be "large businesses" (i.e., the units with 300 and 600 employees). This cutoff value falls into the range of typical definitions for the size of small and medium-sized enterprises (SMEs) used by international organizations such as the World Bank and United Nations. SMEs are typically considered to be enterprises having fewer than 100-250 employees (although an exact definition does not exist) (65).

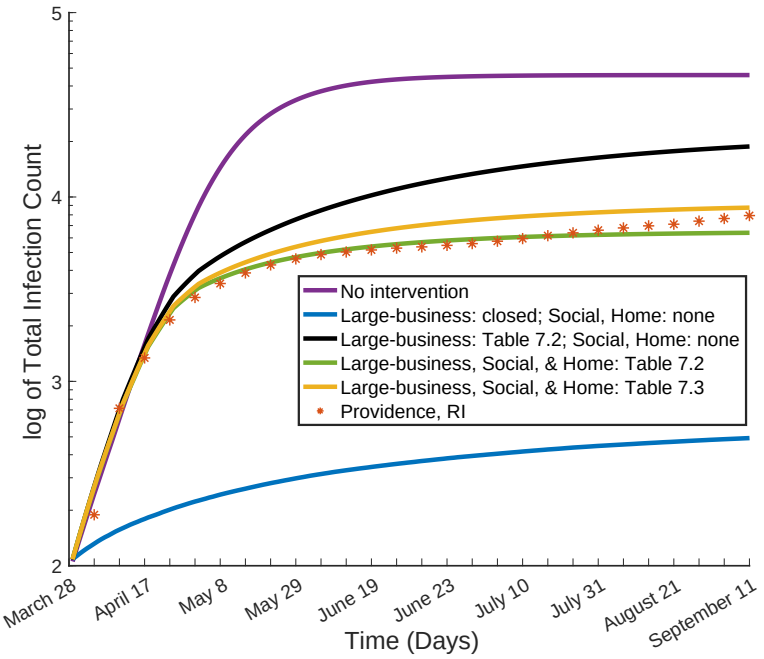

Figure 7.2: Impact of a large-businesses-only closure on the spread of COVID-19. Total infection count over time predicted by simulating the discrete SIR model (6.0.1) on our network for Providence, RI, along with scenarios in which the interactions in large businesses are reduced or removed completely. In all cases, small-business interactions remain unchanged. "Large-business: closed" means that the large-business interactions were removed completely at the start of the intervention. "Social, Home: none" means that no changes were made to the social or home interactions. "Large-business, Social, Home: Table x" means that the large-business, social, and home interactions were reduced according to the scheme described in Table x. The scatter plot represents the weekly total infection count recorded in Providence, RI by the RIDOH (58). We align the date March 28 on the horizontal axis with Day 1 of the intervention in our simulations. Day 1 occurs when the total infection count has surpassed 100 infections. Each of our curves represent the mean over 100 simulations.

We begin by simulating a COVID-19 outbreak on our network for Providence, RI without intervention. Then, when the total infection count has surpassed 100 total infections, we implement a large-business closure on our network. To model a scenario in which large businesses are closed while small businesses remain open, we only lower the frequency of work interactions between individuals in large businesses. In the case where large businesses close completely, we reduce the large-business frequencies to zero.

Percent reductions on interaction frequencies throughout network			
	Social	Large-Business	Home
Week 1 (Day 1)	−5%	−5%	+10%
Week 2 (Day 8)	−10%	−5%	+10%
Week 3 (Day 15)	−15%	−20%	+20%
Week 4 (Day 22)	−25%	−35%	+20%
Week 5 (Day 29)	−35%	−50%	+5%
Week 6 (Day 36)	−25%	−50%	+5%

Table 7.2: Weekly changes in interaction adapted from the values which reproduce the total infection count in Providence, RI during the lockdown in 2020 (58). In this scenario, we keep small businesses open and only lower the work interactions in large businesses. Percentages represent proportional changes in interaction made at the start of each week for all individuals (e.g., "Social: −5%" means all individuals will remove 5% of their current social interactions). "Day 1" of the intervention corresponds to the day that the total infection count surpasses 100 infections. Reductions are applied cumulatively. After Week 6, the resulting interaction frequencies are maintained for the remainder of the simulation.

A complete closure of all large businesses drastically slows the spread of disease: Figure 7.2 shows the result of various closure schemes which apply only to the large businesses. Completely closing the large businesses has the greatest impact on disease dynamics, even when no other changes are made to the interactions between individuals. Here, we consider a scenario in which all large businesses are completely closed on Day 1 of the intervention, but small businesses remain open and social interactions continue as usual. This closure scheme significantly slows the spread of disease during the time it is implemented (blue curve, Figure 7.2). Moreover, we find that the infection rate in our simulated intervention is much lower than the infection rate in Providence, RI during the six-week lockdown (58) which closed all non-essential businesses (labeled March 28 through May 8 in Figure 7.2). After 200 days, the total infection count continues to increase; however, this depressed infection rate remains. Therefore, closing large businesses can act as a way to slow the spread of disease. This could be especially beneficial in scenarios where it is important to reduce the number of individuals infected at a given time.

Percent reductions on interaction frequencies throughout network			
	Social	Large-Business	Home
Week 1 (Day 1)	−2.5%	−5%	+5%
Week 2 (Day 8)	−5%	−5%	+5%
Week 3 (Day 15)	−7.5%	−20%	+10%
Week 4 (Day 22)	−12.5%	−35%	+10%
Week 5 (Day 29)	−17.5%	−50%	+5%
Week 6 (Day 36)	−12.5%	−50%	+5%

Table 7.3: Weekly changes in interaction adapted from the values which reproduce the total infection count in Providence, RI during the lockdown in 2020 (58). In this scenario, we keep small businesses open and only lower the social interactions by half of the values used to reproduce the total infection curve during the lockdown period in Providence, RI. Percentages represent proportional changes in interaction made at the start of each week for all individuals (e.g., "Social: −2.5%" means all individuals will remove 2.5% of their current social interactions). "Day 1" of the intervention corresponds to the day that the total infection count surpasses 100 infections. Reductions are applied cumulatively. After Week 6, the resulting interaction frequencies are maintained for the remainder of the simulation.

The total infection count is not greatly affected by closing small businesses: We also considered a hypothetical scenario in which the lockdown in Providence, RI was applied only to large businesses and social interactions. To model this, we lower the interactions in large businesses by the same amount that interactions were lowered across all businesses in Providence, RI (see Table 7.2), and then maintain the usual interactions in the small businesses (green curve, Figure 7.2). Comparing the green curve in Figure 7.2 to the "Week 6" curve in Figure 7.1a, we find that the total infection count is nearly identical whether we lower the interactions in the small businesses or keep the interactions as usual. In fact, keeping small businesses open in this scenario has essentially no effect on the number of individuals infected or the rate at which these infections occur.

It is worth noting that small businesses often effect social interactions, and therefore a scenario in which small businesses remain open would likely have more social interaction than was observed during the lockdown period in Providence, RI. For

this reason, we tested another scenario in which the social interactions were only lowered by half of the amount that social interactions were lowered during the lockdown period (i.e., the "Social" column in Table 7.1). These values are listed in Table 7.3 and the resulting total infection count is plotted in Figure 7.2 (yellow curve). We find that even when small businesses remain open and social interaction is increased in this way, the total infection count is only slightly greater than the infection data recorded during the lockdown period in Providence, RI (58). In fact, we can substantially slow the spread of disease by lowering the interactions in the large businesses by the same amount as during the lockdown period (i.e., the "Large-Business" column in Table 7.2) and making no changes to the social interactions whatsoever (black curve, Figure 7.2). Thus the significant decrease in interactions within large businesses is a key factor for achieving the lowered infection rate that was observed in Providence, RI during the lockdown period (58).

7.3 Contact Tracing Does Not Reduce the Total Infection Count

In Sections 7.1 and 7.2, we explored intervention strategies which preemptively reduce the interactions of all individuals in the community to control the spread of disease. Now, we consider a contrasting intervention strategy where the interactions of susceptible individuals are only reduced when there is a chance that they have been infected. In this method, after an individual is found to be infected, public health officials "trace," or find, all other individuals who have recently been in close contact with the infected individual. For COVID-19, the CDC designates a close contact to be any individual who has maintained in-person interaction with

an infected individual during the period beginning 2 days before symptom onset (or 2 days before a test was administered) until the infected individual is isolated (66). Then, the individuals considered to be close contacts are asked to voluntarily quarantine for the duration of the disease incubation period. This effort aims to reduce the spread of disease by quickly finding and isolating potential cases, thereby minimizing the chance of exposure for other individuals. For brevity, we refer to this process of widespread contact tracing and subsequent self-quarantine as simply **contact tracing**.

We consider an outbreak of COVID-19 on our network for Providence, RI in which widespread contact tracing is instituted early in the outbreak. Because of the interaction-based nature of our network for Providence, RI, we are able to easily locate the contacts of each newly infected individual. However, our network does not account for the timing of interactions, so we cannot specify a newly infected individual's recent contacts. Instead, we set a threshold frequency $\hat{f}$ (called the **close-contact threshold**) and quarantine any susceptible contact whose total frequency of interaction with the infected individual in all contexts is greater than or equal to $\hat{f}$. That is, for each newly infected individual I_j, we quarantine all susceptible individuals S_i such that $f_{ij} \geq \hat{f}$, where f_{ij} is the sum of interaction frequencies $1/14 \leq f_{ij}^c \leq 14/14$ between S_i and I_j in each interaction context c. We call this collection of susceptible individuals $\{S_i\}_i$ the **close contacts** of I_j. We then model the quarantine by lowering the interaction frequencies of S_i throughout the network.

We mention that this model for contact tracing does not allow us to be precise with the individuals we quarantine. For example, a newly infected individual may not interact with all of its high-frequency contacts during the 2-3 days leading up to a diagnosis, yet our method will quarantine them anyway. Similarly, a newly infected individual may have interacted with some of their low-frequency contacts

during the infectious period, but we do not quarantine any of them. Even so, quarantining an infected individual's high-frequency contacts is a reasonable assumption for our network since high-frequency contacts likely represent housemates, those in a common social circle, and colleagues, which are individuals that are likely to be seen regularly.

We begin by simulating a COVID-19 outbreak on our network for Providence, RI without intervention. Then, matching the timeline of the interventions studied in Section 7.1 and 7.2, we start the contact tracing process when the total infection count has surpassed 100 infections. We account for the pre-symptomatic infectious period for COVID-19 by reducing the interactions of each newly infected individual and the interactions of their close contacts 2-3 days after they become infected (44). For our model, we consider $\hat{f} = 2/14$ to be a reasonable threshold for determining which individuals should be considered close contacts. We note that the average nonzero summed frequency f_{ij} for each pair of individuals i and j in the network is 1.9144/14. For comparison, we also tested several contact tracing schemes when $\hat{f} = 3/14$ and found that the resultant total infection curves were nearly identical to those generated for the same scenarios when $\hat{f} = 2/14$ (results not shown).

Quarantining individuals who may have come in contact with an infected individual does not substantially alter the number of people infected during a COVID-19 outbreak: In Figure 7.3, we plot the impact of several contact tracing schemes which are implemented during a COVID-19 outbreak on our network for Providence, RI. We find that in all realistic scenarios, isolating infected individuals and quarantining their close contacts is not sufficient to lower the number of individuals infected over the epidemic lifespan. Moreover, the rate at which new infections occur in the community is not substantially decreased even when large portions of the infected individuals have their contact traced. Incidentally, we only predict a considerable

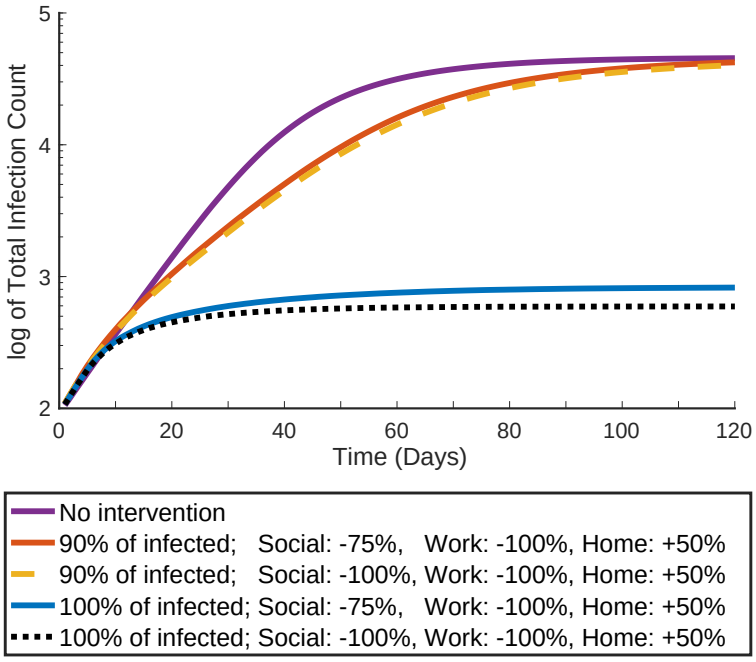

Figure 7.3: Inefficiency of contact tracing in the case of COVID-19. Total infection count over time predicted by simulating the discrete SIR model (6.0.1) on our network for Providence, RI, along with scenarios in which contact tracing is implemented to various extents. In each contact tracing scenario, the close-contact threshold is $\hat{f} = 2/14$. "x% of infected" indicates that x% of all newly infected individuals are isolated and have their close contacts traced and quarantined. The subsequent percentages represent proportional changes in interaction made by the infected individual and their close contacts (e.g., "Social: -75%" means the individuals will remove 75% of their current social interactions). We begin the plot on Day 1 of the intervention in each of our simulations. Day 1 occurs when the total infection count has surpassed 100 infections. Each of our curves represent the mean over 100 simulations.

reduction in the spread of disease when contact tracing is performed for 100% of infected individuals (blue and dotted curves, Figure 7.3). However, this value is not attainable in a real-world scenario since a portion of infections may remain undiscovered and many close contacts may not self-quarantine.

This result aligns with several studies in the literature which found that a large proportion of close contacts must be quarantined before symptom onset to significantly reduce the transmission of SARS-CoV-2 (67, 68, 69). This conclusion depends highly on the pre-symptomatic infectious period of COVID-19. As the length of this

period increases, a higher percentage of transmission occurs before symptom onset. Accordingly, contact tracing must be increased to successfully curtail the epidemic (68, 69). Our model indicates that a 2-3 day pre-symptomatic infectious period in the case of COVID-19 (44) is sufficient to maintain the epidemic even when a large proportion of infected individuals have their contacts traced and quarantined.

Chapter Eight

Discussions and Future Work

The study of infectious diseases and the measures used to control them depend greatly on the underlying interaction structure in the community. However, knowledge of human interactions at the individual level is difficult to obtain given the challenges imposed by large-scale data collection (3). In this work, we propose a method for extending and completing data from the groundbreaking casual-contact survey (1) in a context-specific manner. Our methods, detailed in Chapter 3 and then adapted to the real world in Chapter 5, expand on the approach taken in (1) by constructing unique sub-networks to model the interactions in home, social, and work/school settings. We are able to capture context-specific details which differentiate the interaction structure in each setting, while also maintaining key features of the diary-data (1). In particular, our extended sub-networks reflect the specific degree and cluster distributions revealed in (1) and use the interaction frequencies in this data (1) as weights for our network edges.

We first tested our extended network by simulating the spread of influenza using the discrete SIR model (4.0.1) in Chapter 4. Our results show that the classical differential equation SIR system with the same choice of influenza parameters is unable to reproduce the epidemic size results of the discrete SIR model (Figure 4.1). This suggests that accounting for network structure is crucial for understanding real-world disease transmission. Our network model also predicts that the home, social, and work sub-networks have significantly different effects on epidemic dynamics (Figure 4.2). In particular, we find that while social interactions are less frequent and account for a small percent of infections, they greatly facilitate disease spread by providing connections between work and home clusters.

Realistically, individuals often choose to reduce various interactions after contracting an infectious disease. Accounting for such dynamic responses in our discrete SIR model yields predictions of epidemic-size behavior that are similar to epidemic

data (2) for recent flu seasons (Figure 4.4). One limitation of our model is that it cannot recover the double-peak epidemic size displayed in several flu seasons, as this would likely require knowledge of how the dynamic response to the disease varies with time and epidemic size (12).

Regarding the limitations of our approach, it is important to note that we built our network using a small survey (1) that was collected in a specific location (the University of Warwick). Hence we began by constructing a small, hypothetical population of 3000 individuals, and we studied disease transmission on this network. This strategy allowed us to develop various methods for extending diary-based data in a context-specific way, but our small network cannot fully capture the interaction structure in a larger, more general community.

Later, in Chapter 5, we constructed a larger network of $100,000$ individuals based on the real-world structure of human behavior in Providence, RI. Using household-size data from the United States (34) and businesses-size data from Rhode Island (35), we assigned edges between nodes in the network to represent realistic interactions in the community. Then we adapted the context-specific algorithms from Chapter 3 to inform the frequency of interaction in each context.

We utilized this larger, context-specific interaction network for Providence, RI to study the impact of several intervention strategies introduced early in a COVID-19 outbreak in Chapter 7. Similar to our study of influenza dynamics in Chapter 4, we simulated the spread of COVID-19 on our network using the discrete SIR model (6.0.1). We were able to reproduce the infection data recorded in Providence, RI during the COVID-19 pandemic (58) by greatly reducing the interactions in social and work settings (Figure 7.1). In fact, our results indicate that a decrease in work interactions is necessary to reproduce the lowered infection rate that was observed

during that time. However, we found that the influence of interactions at work on disease progression varied widely by business size. In particular, decreases in the work interactions which occur in small businesses did not significantly affect the number of infections caused by the disease. Yet we can greatly slow the rate of transmission throughout the network by decreasing the interactions which occur in large businesses (Figure 7.2).

Another common strategy for curtailing disease spread is widespread contact tracing and subsequent self-quarantine. The goal in this scenario is to isolate individuals who may have been exposed before they become contagious. However, our findings show that a contact tracing strategy alone does not greatly reduce the total number of individuals infected by SARS-CoV-2 (Figure 7.3). The 2-3 day pre-symptomatic infectious period (44) is sufficient to maintain disease spread in the community.

Our results in Chapters 6 and 7 on the dynamics of COVID-19 are limited in a few ways. First, the available infection data during the COVID-19 pandemic in Providence, RI is actually a measure of positive test results (58). Hence our discrete SIR model is not directly calibrated to the true total infections in Providence, RI. Given more accurate infection data for Providence, RI during the COVID-19 pandemic, we could re-calibrate our model using the same method in Section 6.1 to obtain a better estimate of the transmission rate. Second, the COVID-19 pandemic is still developing and the duration of conferred immunity to SARS-CoV-2 after recovering from the infection is not well-understood. However, following a study in which 95% of test subjects retained immune memory to SARS-CoV-2 after approximately 6 months (38), we assumed that immunity lasts at least 200 days in all recovered individuals so that the susceptible-infected-recovered (SIR) model framework was suitable (27). Given more information on the duration of conferred immunity to SARS-CoV-2, we may find it more appropriate to use an SIRS model, which accounts for short-term

immunity and then a return to susceptibility.

In this work, we take a minimal approach to simulating disease transmission on our extended network, as our focus is on illustrating methods for lifting interaction-based diary data to larger networks. In the future, it may also be interesting to incorporate more realistic dynamic responses to disease onset into our modeling framework to further test our extended network. One could also explore the effects of seasonality on influenza dynamics in different interaction contexts on our network.

The discrete epidemic spread model can also be applied to diseases that do not confer long-term immunity (such as bacterial meningitis). The spread of these diseases is simulated using the susceptible–infected–susceptible (SIS) model, and the probability of transmission is defined as in (4.0.1). Our results for meningitis show that the epidemic size reaches an equilibrium after about 150–250 simulated days, and that the dynamic evolution compares well to meningitis outbreak data from the World Health Organization (70). Similar to our predictions for influenza, we find that changes in the interaction behavior of individuals lead to significant reductions in the peak epidemic size (results not shown). In this work, we considered simple models, such as the SIR and SIS models, to reduce the number of parameters and interactions in our calculations. It would be interesting to consider models such as SEIR (which have an added exposed compartment) within this framework and explore the influence of more complex interactions.

A small percentage of the interactions recorded in the Warwick data (1) occurred in the context of shopping and travel. We also tested the discrete SIR model on networks that included these interactions, which are more likely to take place during weekends, as suggested in the Warwick data. Our extension of these sub-networks to a larger population is based on sampling from the travel and shop degree distri-

butions, as well as specifying all-to-all coupling of nodes in clusters (based on the idea that groups of individuals traveling or shopping together are small and fully clustered). The epidemic size over time given these complete networks is almost identical to the full network results in Figure 4.1 (results not shown), so we chose not to include these contexts in our main results. However, it would be interesting to consider these sub-networks in future simulations of disease spread across several communities generated as in Chapter 5. This approach could be used to study the speed of disease spread across cities, as well as to identify and test strategies for isolating an epidemic. Furthermore, the approach that we proposed for lifting and extending diary-based data in a context-specific way could be extended by differentiating between proximity of interactions in the Warwick data (1) (casual or skin-to-skin). This would allow for a comparison of diseases that spread through casual interactions with those that require close contact between individuals. Moreover, since the extended networks in (1) include interaction proximity, incorporating the type of contact into our networks in the future would offer a more direct means of comparing our results to the conclusions in (1). This could give additional insight into how model predictions depend on the particular methods used to lift diary-based data to extended networks.

Bibliography

[1] Read JM, Eames KT, Edmunds WJ. Dynamic social networks and the implications for the spread of infectious disease. Journal of The Royal Society Interface. 2008;5(26):1001–1007.

[2] Public Health England (PHE). Surveillance of influenza and other respiratory viruses in the UK;. OGL License: https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/. Last accessed 3 September 2019. Available from: https://www.gov.uk/government/statistics/annual-flu-reports.

[3] Keeling MJ, Eames KT. Networks and epidemic models. Journal of the Royal Society Interface. 2005;2(4):295–307.

[4] Zheng X, Zhong Y, Zeng D, Wang FY. Social influence and spread dynamics in social networks. Frontiers of Computer Science. 2012;6(5):611–620.

[5] House T, Ross JV, Sirl D. How big is an outbreak likely to be? Methods for epidemic final-size calculation. In: Proceedings of the Royal Society of London A. vol. 469. The Royal Society; 2013. p. 20120436.

[6] House T. Modelling epidemics on networks. Contemporary Physics. 2012;53(3):213–225.

[7] Tao Z, Zhongqian F, Binghong W. Epidemic dynamics on complex networks. Progress in Natural Science. 2006;16(5):452–457.

[8] Yang Y, Atkinson P, Ettema D. Individual space-time activity-based modelling of infectious disease transmission within a city. Journal of the Royal Society Interface. 2008;5(24):759–772.

[9] Bian L, Huang Y, Mao L, Lim E, Lee G, Yang Y, et al. Modeling individual vulnerability to communicable diseases: A framework and design. Annals of the Association of American Geographers. 2012;102(5):1016–1025.

[10] Burke DS, Epstein JM, Cummings DAT, Parker JI, Cline KC, Singa RM, et al. Individual-based computational modeling of smallpox epidemic control strategies. Academic Emergency Medicine. 2006;13(11):1142–1149.

[11] Eubank S, Guclu H, Kumar VA, Marathe MV, Srinivasan A, Toroczkai Z, et al. Modelling disease outbreaks in realistic urban social networks. Nature. 2004;429(6988):180.

[12] Guo D, Li KC, Peters TR, Snively BM, Poehling KA, Zhou X. Multi-scale modeling for the transmission of influenza and the evaluation of interventions toward it. Scientific Reports. 2015;5.

[13] Riley S. Large-scale spatial-transmission models of infectious disease. Science. 2007;316(5829):1298–1301.

[14] Klovdah A, Potterat J, Woodhouse D, Muth J, Muth S, Darrow W. Social networks and infectious disease: the Colorado Springs Study. Social Science & Medicine. 1994;38(1):79–88.

[15] Liljeros F, Edling C, Amaral L, Stanley H, Aberg Y. The web of human sexual contacts. Nature. 2001;411:907–908.

[16] Wylie J, Jolly A. Patterns of chlamydia and gonorrhea infection in sexual networks in Manitoba, Canada. Sexually Transmitted Diseases. 2001;28(1):14–24.

[17] Edmunds W, Kafatos G, Wallinga J, Mossong J. Mixing patterns and the spread of close-contact infectious diseases. Emerging Themes in Epidemiology. 2006;3(10).

[18] Newman M. The Structure and Function of Complex Networks. SIAM Review. 2003;45(2):167–256.

[19] Mallory K, Abrams JR, Schwartz A, Ciocanel MV, Volkening A, Sandstede B. Influenza spread on context-specific networks lifted from interaction-based diary data. Royal Society Open Science. 2021;8(1):191876.

[20] Hardiman SJ, Katzir L. Estimating clustering coefficients and size of social networks via random walk. In: Proceedings of the 22nd international conference on World Wide Web. International World Wide Web Conferences Steering Committee; 2013. p. 539–550.

[21] Ahn YY, Han S, Kwak H, Moon S, Jeong H. Analysis of topological characteristics of huge online social networking services. In: Proceedings of the 16th international conference on World Wide Web. ACM; 2007. p. 835–844.

[22] Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature. 1998;393(6684):440–442.

[23] Albert R, Barabási AL. Statistical mechanics of complex networks. Reviews of Modern Physics. 2002;74(1):47.

[24] Barabási AL, Albert R. Emergence of scaling in random networks. Science. 1999;286(5439):509–512.

[25] Bianconi G, Barabási AL. Competition and multiscaling in evolving networks. EPL (Europhysics Letters). 2001;54(4):436.

[26] Inter-Departmental Business Register. UK Business: Activity, size and location: 2015;. Last accessed 3 September 2019. Available from: https://www.ons.gov.uk/businessindustryandtrade/business/activitysizeandlocation/bulletins/ukbusinessactivitysizeandlocation/2015-10-06.

[27] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. In: Proceedings of the Royal Society of London A: Mathematical, physical and engineering sciences. vol. 115. The Royal Society; 1927. p. 700–721.

[28] Longini IM, Halloran ME, Nizam A, Yang Y. Containing pandemic influenza with antiviral agents. American Journal of Epidemiology. 2004;159(7):623–633.

[29] Tuite AR, Greer AL, Whelan M, Winter AL, Lee B, Yan P, et al. Estimated epidemiologic parameters and morbidity associated with pandemic H1N1 influenza. Canadian Medical Association Journal. 2010;182(2):131–136.

[30] Zhang S. Estimating transmissibility of seasonal influenza virus by surveillance data. Journal of Data Science. 2011;9:55–64.

[31] Ellner SP, Guckenheimer J. Dynamic models in biology. Princeton University Press; 2011.

[32] Rohatgi A. WebPlotDigitizer;. Version 4.2. Last accessed 3 September 2019. Available from: https://automeris.io/WebPlotDigitizer.

[33] US Census Bureau, Population Division. Annual Estimates of the Resident Population for Incorporated Places of 50,000 or More, Ranked by July 1, 2019 Population: April 1, 2010 to July 1, 2019 (SUB-IP-EST2019-ANNRNK);. Last accessed 24 February 2021. Available from: https://www.census.gov/data/tables/time-series/demo/popest/2010s-total-cities-and-towns.html#tables.

[34] United States Census Bureau. Historical Households Tables: Table HH-4. Households by Size: 1960 to Present;. Last accessed 24 February 2021. Available from: https://www.census.gov/data/tables/time-series/demo/families/households.html.

[35] United States Census Bureau. 2017 SUSB Annual Data Tables by Establishment Industry: Number of Firms, Number of Establishments, Employment, and Annual Payroll by Enterprise Employment Size for the United States and States, Totals: 2017;. Last accessed 24 February 2021. Available from: https://www.census.gov/data/tables/2017/econ/susb/2017-susb-annual.html.

[36] Zhang R, Li Y, Zhang AL, Wang Y, Molina MJ. Identifying airborne transmission as the dominant route for the spread of COVID-19. Proceedings of the National Academy of Sciences of the United States of America. 2020;117(26):14857–14863.

[37] Seungjae L, Tark K, Eunjung L, et al. Clinical course and molecular viral shedding among asymptomatic and symptomatic patients with SARS-CoV-2 infection in a community treatment center in the Republic of Korea. JAMA Internal Medicine. 2020;180(11):1447–1452.

[38] Dan JM, Mateus J, Kato Y, Hastie KM, Yu ED, Faliti CE, et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. Science. 2021;371(6529).

[39] Bullard J, Dust K, Funk D, Strong JE, Alexander D, Garnett L, et al. Predicting Infectious Severe Acute Respiratory Syndrome Coronavirus 2 From Diagnostic Samples. Clinical Infectious Diseases. 2020;71(10):2663–2666.

[40] Young BE, Ong SWX, Ng LF, Anderson DE, Chia WN, Chia PY, et al.. Immunological and Viral Correlates of COVID-19 Disease Severity: A Prospective Cohort Study of the First 100 Patients in Singapore;. Last accessed 24 February 2021. Available from: http://dx.doi.org/10.2139/ssrn.3576846.

[41] Wölfel R, Corman VM, Guggemos W, et al. Virological assessment of hospitalized patients with COVID-2019. Nature. 2020;581:465—469.

[42] SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. In: The Lancet Microbe, ISSN: 2666–5247. vol. 2, issue: 1; 2021. p. e13–e22.

[43] van Kampen JJA, van de Vijver DAMC, Fraaij PLA, et al. Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (COVID-19). Nature Communications. 2021;12(267).

[44] He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nature Medicine. 2020;26:672–675.

[45] Dietz K. The estimation of the basic reproduction number for infectious diseases. Statistical Methods in Medical Research. 1993;2(1):23–41.

[46] Centers for Disease Control and Prevention (CDC). COVID-19 Pandemic Planning Scenarios;. Last accessed 24 February 2021. Available from: https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html.

[47] Chinazzi M, Davis JT, Ajelli M, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science. 2020;368(6489):395–400.

[48] Imai N, Cori A, Dorigatti I, Baguelin M, Donnelly CA, Riley S, et al.. Report 3: Transmissibility of 2019-nCoV;. Last accessed 24 February 2021. Available from: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-transmissibility-25-01-2020.pdf.

[49] Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. The New England Journal of Medicine. 2020;382(13):1199–1207.

[50] Munayco CV, Tariq A, Rothenberg R, et al. Early transmission dynamics of COVID-19 in a southern hemisphere setting: Lima-Peru: February 29th-March 30th, 2020. Infectious Disease Modelling. 2020;5:338–345.

[51] Salje H, Tran Kiem C, Lefrancq N, et al. Estimating the burden of SARS-CoV-2 in France. Science. 2020;369(6500):208–211.

[52] Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the Basic Reproduction Number (R0). Emerging Infectious Diseases. 2019;25(1):1–4.

[53] Al-Raeei M. The basic reproduction number of the new coronavirus pandemic with mortality for India, the Syrian Arab Republic, the United States, Yemen, China, France, Nigeria and Russia with different rate of cases. Clinical Epidemiology and Global Health. 2021;9:147–149.

[54] Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. International Journal of Infectious Diseases. 2020;92:214–217.

[55] Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet. 2020;395(10225):689–697.

[56] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. The New England Journal of Medicine. 2020;382(13):1199–1207.

[57] Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Eurosurveillance. 2020;25(4):2000058.

[58] Rhode Island Derpartment of Health. Rhode Island COVID-19 Response Data;. Last accessed 12 January 2021. Available from: ri-department-of-health-covid-19-data-rihealth.hub.arcgis.com.

[59] Raimondo GM. Executive Order 20-14: March 28, 2020. Amended Eleventh Supplemental Emergency Declaration – Staying at Home, Reducing Gatherings, Certain Retail Business Closures and Further Quarantine Provisions;. Last accessed 24 February 2021. Available from: https://governor.ri.gov/documents/orders/Executive-Order-20-14.pdf.

[60] Raimondo GM. Executive Order 20-32: May 8, 2020. Twenty-ninth Supplemental Emergency Declaration – Lifting Stay at Home Order, Leaving Certain Restrictions in Place, Limited Re-Openings with Stringent Health Regulations;. Last accessed 24 February 2021. Available from: https://governor.ri.gov/documents/orders/Executive-Order-20-32.pdf.

[61] Das A, Dhar A, Goyal S, Kundu A, Pandey S. COVID-19: Analytic results for a modified SEIR model and comparison of different intervention strategies. Chaos, Solitons & Fractals. 2021;144:110595.

[62] Glass DH. European and US lockdowns and second waves during the COVID-19 pandemic. Mathematical Biosciences. 2020;330:108472.

[63] Lalwani S, Sahni G, Mewara B, Kumar R. Predicting optimal lockdown period with parametric approach using three-phase maturation SIRD model for COVID-19 pandemic. Chaos, Solitons & Fractals. 2020;138:109939.

[64] Gevertz JL, Greene JM, Sanchez-Tapia CH, Sontag ED. A novel COVID-19 epidemiological model with explicit susceptible and asymptomatic isolation compartments reveals unexpected consequences of timing social distancing. Journal of Theoretical Biology. 2021;510:110539.

[65] World Bank Group. World Bank Group Support for Small and Medium Enterprises (SMEs) – 2019 Report;. Last accessed 24 February 2021. Available from: https://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/SME_Synthesis.pdf.

[66] Centers for Disease Control and Prevention (CDC). Interim Guidance on Developing a COVID-19 Case Investigation Contact Tracing Plan;. Last accessed 24 February 2021. Available from: https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/overview.html.

[67] Peak CM, Kahn R, Grad YH, Childs LM, Li R, Lipsitch M, et al. Individual quarantine versus active monitoring of contacts for the mitigation of COVID-19: a modelling study. The Lancet Infectious Diseases. 2020;20(9):1025–1033.

[68] Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020;368:6491.

[69] Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. The Lancet Global Health. 2020;8(4):e488–e496.

[70] World Health Organization (WHO). WHO Multi-Disease Surveillance Centre Ouagadougou, Regional Meningitis Surveillance;. Last accessed 30 September 2016. Available from: http://www.who.int/csr/disease/meningococcal/epidemiological/en/.