

DESIGNING GOOD POLICIES FOR MEDICAL APPLICATIONS

A Dissertation
Presented to
The Academic Faculty

By

Timur Tankayev

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology

May 2020

Copyright © Timur Tankayev 2020

DESIGNING GOOD POLICIES FOR MEDICAL APPLICATIONS

Approved by:

Dr. Craig Tovey, Advisor
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Dr. Christos Alexopoulos
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Dr. David Goldman
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Dr. Paul Griffin
School of Industrial Engineering
Purdue University

Dr. Turgay Ayer
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Date Approved: March 14, 2020

To my grandparents.

ACKNOWLEDGEMENTS

I am indebted to many wonderful people who have supported me through the last few years.

First of all, I want to thank my advisor Craig Tovey. Without his support and guidance this work would have been impossible. You have not only been my academic advisor but also an incredible mentor.

I want to thank my committee Members: Christos Alexopoulos, David Goldsman, Paul Griffin, and Turgay Ayer for generously devoting their valuable time to serve on my committee. Thank you for the helpful discussions and feedback.

I would like to thank wonderful faculty at Georgia Tech. In particular, professors Santanu Dey, Robert Foley, David Goldberg, Renato Monteiro, Martin Savelsbergh, and Yajun Mei. Your classes had a significant impact in developing my intellectual curiosity. I would also like to thank wonderful administrative staff at ISyE, especially Dawn Strickland and Amanda Ford. You have made my academic life as smooth and trouble-free as I could have hoped for.

I am grateful to the friends I have met at Georgia Tech: Luke Marshall, Mina Georgieva, Giorgio Morbioli, Tony Yaacoub, Idil Arsik, Seyma Guven-Kocak, Asteroide Santana, and many more.

Finally, I want to thank my family, my parents, aunts, and cousins, for all their support.

TABLE OF CONTENTS

Acknowledgments	iv
List of Tables	viii
List of Figures	ix
Chapter 1: Introduction and Background	1
1.1 Ranking and Selection	1
1.2 Epidemiology	2
1.3 Contributions	5
I Multinomial Selection	7
Chapter 2: Optimal Solution to the Multinomial Selection Problem for Two Alternatives	8
2.1 Introduction	8
2.2 Definitions and elementary properties	10
2.3 Holey and unholey policies	15
2.4 Slipping to optimality	22
2.5 Conclusions	28
II Epidemiology	30
Chapter 3: Basic Reproduction Number: Theoretical Analysis	31

3.1	Introduction	31
3.2	Theoretical Analysis	32
3.3	Conclusions	39
Chapter 4: Case study of the Yemen cholera epidemic		40
4.1	Introduction	40
4.2	Case study	40
4.2.1	Model overview	41
4.2.2	Results	42
4.3	Cholera Outbreak Model Specifications	44
4.3.1	Contact network	45
4.3.2	Disease progression and transmission	45
4.3.3	Simulation	46
4.3.4	Model limitations and validation	47
4.4	Conclusions	49
Chapter 5: Epidemic Control		52
5.1	Introduction	52
5.2	Problem statement	56
5.3	Hardness and inapproximability	59
5.3.1	Approximability	62
5.3.2	Discussion and practical performance	65
5.4	Experiments	66
5.5	Conclusions	72

Appendix A: The LP Formulation	75
Appendix B: Counterexample to a stronger statement than Lemma 2.3.1	76
Appendix C: Proofs of Theorems in Chapter 3	79
Appendix D: Experimental results for the optimal intervention	83
References	101

LIST OF TABLES

4.1	Cholera outbreaks, their R_0 and affected population	44
4.2	Simulation parameters	47
B.1	PCS and expected number of trials for deterministic policies (part 1) . .	77
B.2	PCS and expected number of trials for deterministic policies (part 2) . .	78
D.1	Karate network intervention results. No intervention, SAAIP, and R_0 based interventions	84
D.2	Karate network intervention results continued. Interventions based on betweenness and eigenvector centralities.	85
D.3	Kenyan village network intervention results. No intervention, SAAIP, and R_0 based interventions	86
D.4	Kenyan village network intervention results continued. Interventions based on betweenness and eigenvector centralities.	87
D.5	French high school network intervention results. No intervention, SAAIP, and R_0 based interventions	88
D.6	French high school network intervention results continued. Interven- tions based on betweenness and eigenvector centralities.	89

LIST OF FIGURES

2.1	Example of a policy. All red states are terminal. All white states are traversable. Unreachable states are not shown. \mathbf{u} and \mathbf{v} are terminal but not holes as they have untraversable successors in at least one direction. w is a hole as both $\mathbf{w} + (1, 0)$ and $\mathbf{w} + (0, 1)$ are traversable.	16
2.2	Policy that terminates if an alternative has four wins but also terminates at $\mathbf{w} : [\mathbf{w}] = (1, 2)$.	17
2.3	Two possibilities for terminal states to block \mathbf{u} from changing its decision. At least one of the green or blue sets of dots must be terminal as $[\mathbf{v}]$ is not a hole. On the left $w_1 < v_1$ and $w_2 > v_2$. On the right $w_1 = v_1$ and $v_2 < w_2$	22
2.4	The inductive step in Lemma 2.4.1	24
3.1	Construction of a graph that has higher R_0 than any given graph G, but much lower F. R_0 has increased, as for unchanged duplicated vertices it stays the same and we added an edge. F has decreased as the spread to G' is conditioned on traversing the edge (u, u')	36
3.2	Example of a real population distribution along a line. The Russian population in its Asian regions closely follows the Trans-Siberian Railway. The numbers in the legend are densities in units of people per square kilometer ([72]).	37
3.3	Construction of a graph that for arbitrarily low R_0 has constant F. This is a member ($n = 28$) of a sequence of graphs with arbitrarily low fixed R_0 that always have at least F fraction of infected. H is a well-connected central component with constant degree vertices (to keep R_0 from growing). A bounded fraction of blue vertices is added to H to get R_0 as low as required.	38

4.1	Contact network Example of the contact network for a an approximately 4 km ² region of Raymah Governorate. Blue nodes and links represent water sources and their connections. Orange nodes represent people.	42
4.2	Diagram of the cholera model. People can be Susceptible to cholera, Exposed, Infectious or Removed (immune to infection and not part of the epidemic process). Black arrows represent state transitions. Susceptible people get infected by exposure to bacteria in the water. After an incubation period, they transition from Exposed to Infectious state. Cholera bacteria are shed into the water supply by Infectious people. There is a natural rate of bacterial decay that removes them from the environment.	43
4.3	Example of a simulated (left) vs reported (right) disease spread. Large numbers of infections occur in densely populated urban areas around the capital of Sana'a and western parts of the country. Due to the model not accounting for population movements, the simulated epidemic has fewer infections in the main refugee destinations: Hajjah and Amran.	43
4.4	Final epidemic sizes for varying probabilities of distant infections. Increasing probability of distant infection by a factor of 2 results in a significant increase in the final epidemic size but has negligible impact on R_0 . . .	44
4.5	Plot of estimated infections by governorate. The numbers reported by WHO through February 2nd 2018 [75] are compared to model projections (median with 95% CI) for the same period. *Northwestern governorates that were the main refugee destination. **City of Aden and surrounding governorates that experienced heaviest fighting and initial refugee flood. . .	51
5.1	Construction for the Proof of Proposition 5.3.1. s is a unique starting vertex for each scenario in S . Vertices e_i correspond to the edges of the graph in the Clique Problem. Vertex groups v_j correspond to the vertices the Clique Problem	61
5.2	Construction for the Proof of Proposition 5.3.2. s is a unique starting vertex for each scenario in S . Vertices v_1 and v_2 have two successors each. Vertices v_3 and v_4 are connected to n vertices.	63
5.3	Construction for the Proof of Proposition 5.3.3. s is a unique starting vertex for each scenario in S . s is connected to vertices $\{v_1, \dots, v_{b+1}\} = V_1$. Each vertex in V_1 has an arc to each vertex in V_2 . By construction, with a budget b we cannot save more than b vertices.	64

5.4	Karate club , $b = 1$. Expected infections $E[R]$ against probability of transmission p per interaction for the Karate network with intervention budget of 1. The curves (from top to bottom) correspond expected infections coming from intervention policy based on: no intervention, SAAIP, minimizing R_0 , removing node with highest betweenness centrality, and removing node with highest eigenvalue centrality. Note that in this case, R_0 and eigenvalue centrality based interventions produce the same policy. However, the optimal policy is the one based on SAAIP and betweenness centrality interventions.	69
5.5	Karate club , $b = 5$. All interventions except the betweenness centrality produce the same policies.	70
5.6	Karate club , $b = 7$. SAAIP based intervention strictly dominates all other interventions.	70
5.7	Kenyan village , $b = 5$. Again, SAAIP based intervention strictly dominates all other interventions.	71
5.8	Kenyan village , $b = 10$. SAAIP based intervention policy dominates all other policies. However, it becomes different from R_0 based policy for large values of p	71
5.9	French high school , $b = 8$. R_0 based intervention slightly outperforms SAAIP for small values of p . Running SAAIP with larger number of scenarios fixes the issue. Also, notice that for very high values of p all policies perform about the same, but SAAIP is much better for moderate to high probabilities of transmission.	72
B.1	Holey policy	76

SUMMARY

This thesis examines two topics at the intersection of mathematical decision-making and healthcare. The first part addresses a problem of designing optimal policy in ranking and selection. The second part critiques the standard mathematical measure of outbreak severity in epidemiology, proposes a more accurate measure, and discusses how to design an effective prevention policy.

Chapter 2 of this thesis deals with the multinomial selection problem (MSP), a problem in ranking and selection. MSPs arise when designing a protocol to select the most effective drug or treatment from among multiple alternatives. The objective of MSP is to find a stopping policy for repeated independent trials, each of which reports a winner among competing alternatives, that has low expected cost and high probability of correct selection (PCS) of the best alternative. In 1959, Bechhofer, Elmaghraby and Morse formulated the problem as minimizing the worst-case expected number of trials, subject to a lower bound on PCS and upper bound on the maximum number of trials, over all probability vectors outside an indifference zone. For the case of two alternatives, we prove that if one employs a particular probability vector known as the slippage configuration, then a linear program always finds an optimal stopping policy.

Chapters 3 and 4 discuss the basic reproduction number, a standard measure of potential disease spread. A proxy for the computationally intractable expected fraction of the population to be infected, it is intended to be less than one if the outbreak will die out, and to exceed one if the outbreak will become pandemic. It has long been used to predict the urgency and efficacy of proposed interventions by public health organizations which must determine the best use of limited resources. However, traditional homogeneous contact models have been largely replaced by more accurate heterogeneous contact network models. We prove that in shifting to heterogeneous contact models, the reproduction number loses its crucial theoretical properties. It cannot be used to approximate the scale of the

epidemic, does not provide a threshold, and lacks a fundamental monotonicity property. Its worst-case inaccuracy is infinite. We propose to replace the reproduction number by an approximation of the expected fraction of population infected. We prove that accurate approximation is computationally feasible. We conduct a case study of a fine-grained spatial network model of the ongoing cholera outbreak in Yemen. We find that the reproduction number neither aids in assessing severity, nor identifies the important factors that affect the outbreak

One of the main motivations behind studying the spread of diseases is to mitigate their impact. Resource scarcity makes developing good prevention strategies a challenging optimization problem. In Chapter 5, we tackle the problem of minimizing disease spread on a contact network subject to a limited immunization budget. We present a stochastic programming formulation to design an intervention and discuss its computational complexity and approximability. On a more practical side, we evaluate the performance of such intervention against alternatives studied in the literature. We conduct this evaluation on three real contact networks and find that our approach significantly outperforms the alternatives. Moreover, the discrepancy is especially pronounced in the most dangerous cases of highly infectious diseases and tight intervention budgets.

CHAPTER 1

INTRODUCTION AND BACKGROUND

The healthcare industry is one of the largest and fastest growing industries in the world. In 2018, in the US alone, 17.7% of gross domestic product was related to health care spending [1]. This amounts to a staggering figure of almost 3.5 trillion dollars. Improved decision-making in healthcare has the potential to save significant amounts of money [2]. More importantly, it can save and improve people's lives. Optimization/Operations Research already has made solid contributions to medical decision-making. In turn, medical applications have inspired the development of new methodologies and solution techniques [3]. In this thesis, we continue this fruitful collaboration by applying optimization techniques to obtain optimal policies in two areas of public health. In the first part of this thesis, we deal with finding an optimal policy for a problem in ranking and selection with possible applications to clinical trials. In the second part of the thesis, we explore mathematical epidemiology. We critique the standard measure of outbreak severity, provide an alternative, and study optimal targeted prevention policies.

1.1 Ranking and Selection

Ranking and selection procedures are statistical techniques for comparing a set of alternatives, under the assumption that they are not all the same [4]. Sequential selection procedures are often used by the simulation community for output analysis [5]. Their allure comes from the fact that sequential sampling can provide statistical insights in much fewer expensive simulation runs than traditional methods. In a similar vein, our interest comes from the problem of designing clinical trials. Clinical trials are extremely expensive, with a median cost of around 19 million dollars [6]. Furthermore, they can require large sample sizes and can lack power to evaluate drug efficacy [7]. Finally, classical clinical trial

designs can needlessly assign patients to ineffectual treatments, endangering their health [8]. A lot of these issues can be alleviated by pursuing a fully sequential study design. In a sequential approach, the trials can stop as soon as sufficient evidence of comparative efficacy is obtained, thus eliminating unnecessary costs and risks. Recently, there has been a resurgence of such approaches, especially from the angle of very related contextual bandit problems [9, 10] Deciding on optimal stopping point for such problem is the focus of the second chapter of this thesis.

More specifically, we work on the multinomial selection problem, specified precisely in the next chapter. There are many problem setups in ranking and selection with different notions of ‘best’ alternatives, assumptions on the available comparison, and constraints on the selection policy. We chose to work in a very clean setting with minimal assumptions. Informally, alternatives compete in sequential independent trials and we get to observe the winner of the trial. Our goal is to choose the alternative with the highest chance of winning within some given confidence (e.g. at least 95%) in the smallest (expected) number of trials. The problem was posed more than 60 years ago [11] and since then, many stopping policies have been proposed [12, 11, 13, 14]. Recently, [15] constructed a policy obtained by solving a linear program, which has the potential to be truly optimal. [16] proved that for two alternatives, with appropriate inputs, this LP based policy guarantees the required confidence. In this work, we demonstrate that in the same setting it is also truly optimal. I.e. we prove that the policy generated by LP with appropriate parameters does indeed minimize the worst case expected number of trials.

1.2 Epidemiology

The second part of this thesis is devoted to another topic relevant to health policy: epidemiology. Historically, infectious diseases have been the leading sources of human death and suffering [17]. For instance, Spanish flu alone caused about the same number of casualties as both World Wars combined [18]. Humanity has been fighting some infectious

diseases, such as malaria, flu, measles, and tuberculosis for centuries. Meanwhile, there is also a constant threat of newly emerging diseases such as Wuhan Coronavirus, Zika, and Ebola. Given the danger, it is important that our mathematical and computational models of these outbreaks be functional. We address this issue in Chapters 3, 4 and 5. In chapter 3, we examine the basic reproduction number, one of the main measures of the danger of a potential outbreak. We expose its theoretical deficiencies, critique its applicability, and propose a better alternative. In the following chapter, we complement our theoretical findings with a large fine-scale case study of the current cholera epidemic in Yemen. Finally, in Chapter 5, we discuss how to prevent the spread of disease using targeted interventions.

Mathematical models of infectious disease can be traced back all the way to 18th century [19]. Since the beginning of the 20th century, there has been a lot of development in the realm of compartmental models [20, 21]. These models aim to describe the dynamics of the disease as it is transmitted through a population. Perhaps the most famous example is the Kermack-McKendrick differential equation model [20]. The original model can be derived as a limit of a stochastic process and has age stratification. We employ a slightly simplified SIR model to illustrate important concepts. Suppose each member of the population belongs to one of the three compartments: susceptible (S), infectious (I), or removed (R). The SIR process evolves according to the following set of differential equations:

$$\begin{aligned}\frac{dS}{dt} &= -\beta IS \\ \frac{dI}{dt} &= \beta IS - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

, where β is the transmitting contact rate and γ is the rate of recovery, with the initial conditions $I + S + R = 1$ and $I = \epsilon, R = 0$. Note that the dynamics of the system primarily depend on the ratio:

$$R_0 = \frac{\beta}{\gamma}$$

If $R_0 > 1$, then initially $\frac{dI}{dt} > 0$, i.e. the number of infected will increase and there will be an outbreak. Otherwise, if $R_0 < 1$ we have $\frac{dI}{dt} < 0$, i.e. the number of infected decreases and there is no outbreak. This ratio, called the basic reproduction number, has a physical meaning – it can be thought of as the number of direct infections caused by an initially infectious person in an otherwise completely susceptible population.

These old models are a bit too crude and often don't produce useful epidemiological insights [22]. To address this issue, there has been a significant development of more sophisticated models [23]. For instance, newer network models attempt to capture more complex transmission dynamics, stochasticity of the underlying processes, and heterogeneous mixing of the susceptible population. Some authors have reported empirical shortcomings of R_0 in this new setting [24]. Nevertheless, public health organizations still use R_0 to obtain qualitative and quantitative guidance and inform their policy decisions [25, 26, 27]. We attempt to remedy this situation by conclusively demonstrating that it is a poor measure of epidemic spread potential. We explore theoretical properties of R_0 and prove that the assumed threshold property is invalid in network models. Moreover, it can be arbitrarily inaccurate as a predictor of the size of the potential outbreak. Meanwhile, we propose a much more straightforward measure, the expected fraction of population that will get infected, and demonstrate that it can be approximated efficiently to arbitrary accuracy.

The reproduction number's lack of desirable theoretical properties does not rule out the possibility that it might be useful for models that occur in practice. We test this possibility by studying a large-scale ongoing epidemic. We chose a large-scale epidemic to complement previous assessments conducted on small scale outbreaks and networks [28]. One of the largest current infectious disease outbreaks is the ongoing cholera epidemic in Yemen. Since 2016, it caused over a million reported cases and thousands of deaths. We build a very high-resolution network model of Yemeni population and waterway network. The resulting network has over 25 million nodes and 10^{13} edges. After calibrating the model using historical data, we investigate the performance of R_0 as a predictor. We find that

the basic reproduction number neither aids in assessing the severity of the outbreak nor identifies important factors that affect disease spread.

Finally, in the last chapter, we investigate the topic of outbreak prevention. We study how to design a targeted intervention on a network to minimize the expected spread of infection. The approaches in the literature come from many different perspectives and are all heuristic [29]. The most common approaches are based on minimizing various network parameters, for instance, the aforementioned basic reproduction number (or equivalently average vertex degree) [30, 31]. In contrast, our goal is to design a policy that is truly optimal for the objective of minimizing the expected number of infections. To do so, we devise a stochastic programming approach that rapidly converges to the true optimum. We discuss the issues of its theoretical tractability and perform computational experiments on real contact networks. We find that the approach is feasible for small networks. Furthermore, in practice it can significantly outperform the alternatives proposed in the literature.

1.3 Contributions

There are two recurring themes in this thesis. The first one, healthcare, motivates the problems. The second one is the pursuit of optimality. When both lives and large sums of money are at stake, it is important that the solutions be the best possible or at least have a guarantee on their closeness to optimality. It is even more important that optimality be measured against the correct goal. This theme permeates throughout this work. For the multinomial selection problem, we deal with a concrete finite optimal policy, rather than with an unbounded one that is optimal in some asymptotic regime. In epidemiology, we argue for using a correct metric to measure the potential disease spread. For preventive measures, again we focus on targeted intervention that truly minimizes the impact of the disease. Overall, we hope this work helps to demonstrate the importance of proper goals.

In conclusion, our contributions can be summarized as follows. In the context of ranking and selection, we prove that LP based policy is optimal for two alternatives. This

partially resolves a 60-year-old question. In the context of epidemiology, we demonstrate the deficiency of measuring outbreak severity using R_0 , the basic reproduction number. We prove that for network models, an arbitrarily small R_0 can lead to a large disease outbreak. And at the other extreme, an negligible outbreak can have an arbitrarily large R_0 . We offer an alternative to R_0 , namely, the expected fraction of the population to be infected, for which R_0 at best is merely a proxy. We prove that the expected fraction infected is computationally practicable, despite its being $\#P$ -hard to compute exactly. As to practical public health policy, we show that in the case of the cholera epidemic in Yemen, knowing R_0 does not provide any meaningful insights about the scope, size, and danger of the outbreak. Furthermore, R_0 is insensitive to important epidemiological factors, such as long-distance disease transmission. Finally, we develop a policy based on stochastic programming that converges to the truly optimal targeted intervention. We test it on real world contact networks and show that it performs significantly better than the alternatives in the literature.

Part I

Multinomial Selection

CHAPTER 2
OPTIMAL SOLUTION TO THE MULTINOMIAL SELECTION PROBLEM FOR
TWO ALTERNATIVES

2.1 Introduction

The multinomial selection problem (MSP) is to select the best out of $k \geq 2$ competing alternatives. There is an unknown probability vector $\mathbf{p} = (p_1, \dots, p_k) \in \mathcal{P}$, where $\mathcal{P} \equiv \left\{ \mathbf{p} \in (0, 1]^k : \sum_{i=1}^k p_i = 1 \right\}$. The alternatives compete in successive independent trials. Each trial is won by one alternative, with alternative i having probability $p_i > 0$ of winning. Without loss of generality, let $p_1 \leq p_2 \leq \dots \leq p_k$, with the understanding that this ordering of the p_i is not known during the trials. Alternative p_k is called the best, or most probable.

A long-standing research goal of the field is to find a policy that conducts the minimum expected number of trials, subject to identifying the best alternative with high probability and an upper bound on the maximum allowed number of trials. As stated, the problem is poorly posed because, for example with $k = 2$, any finite upper bound u , and any $\delta > \frac{1}{2}$, it is impossible to succeed with probability $\geq \delta$ for the probability vector $p_i = \frac{1}{2} + \epsilon(-1)^i$ for $\epsilon > 0$ sufficiently small. [32] provided a workable formulation of the problem which has become standard. For $\rho > 1$ define the *indifference zone* to be

$$Z_\rho \equiv \left\{ \mathbf{p} \in \mathcal{P}; p_i \leq p_{i+1} \forall i \leq k-1; \frac{p_k}{p_{k-1}} < \rho \right\}.$$

The idea is that ρ is the smallest difference worth detecting. Given $\rho > 1$, Bechhofer's formulation is to minimize the worst-case expected number of trials subject to given lower and upper bounds, respectively, on the PCS and maximum number of trials. The worst case is taken over all probability vectors not in Z_ρ . The principal result of this chapter is the resolution of this problem for $k = 2$.

Numerous policies have been proposed over the past 60 years, beginning with those of [12] and [11], which determine a fixed number of trials for given ρ and PCS bound. [13] and [33] propose dynamic stopping rules, ones that decide whether or not to continue after each trial, based on the trial outcomes so far. [34] develops a dynamic stopping rule based on the difference between the largest and second-largest numbers of wins, which for $k = 2$ is a gambler’s ruin problem. [14] propose the first hybrid policy, a combination of those in [13] and [34]. [35] introduce the idea of *curtailment*, which is to improve on a static policy by stopping when the maximum allowed number of additional trials could not affect the outcome. [36] propose a partially curtailed version of Bechhofer’s original policy. [37] applies additional curtailment to this policy. [38] propose a truncated and curtailed version of the policy in [33] and compute parameter values for it in [39]. [40] refines curtailment and the sampling concept of [13], and later with Hsu develops a hybrid policy in [41]. [42] propose the AVC (all vector comparisons) policy for cases where the winning alternative has a largest numerical measure, which employs bootstrapping to increase efficiency; [43] experimentally compare AVC with other policies. [44] present new policies for both small and large sample sizes. [45] and [46] compare several policies and provide more accurate parameter values than computed previously. [47] propose a novel policy that is designed to be practical when the number of alternatives is potentially very large. [15] formulate a family of linear programming models with the following property: For given k , $\rho > 1$, upper bound u , PCS lower bound ζ , and probability vector $\mathbf{p} \notin Z_\rho$, the optimal solution to a member of the family specifies a policy that minimizes the expected number of trials for the vector \mathbf{p} , subject to the upper and lower bounds, or determines that no feasible stopping rule exists. This method was the first to generate randomized policies, which are easily shown to be necessary for optimality in some cases. For completeness here, we include an equivalent LP model in the Appendix (Section A). An associated integer program in [15] finds the analogous deterministic policy.

The LP formulation of [15] begs the question of which probability vector \mathbf{p} to employ.

The natural candidate for the role of \mathbf{p} is the so-called *slippage configuration*, defined by $\frac{p_k}{p_i} = \rho \forall 1 \leq i \leq k-1$. [48] introduce the slippage configuration and prove it has minimum PCS over all probability vectors not in Z_ρ , for the policy in [11]. The slippage configuration has since been proved to have minimum PCS, over all probability vectors outside the indifference zone Z_ρ , for many specific policies (see the survey in [42]). Recently a general result for $k = 2$ has been obtained. Define a policy to be *sane* if, upon termination, it selects an alternative with the most wins. Then the slippage configuration has minimum PCS (outside Z_ρ) for all sane policies when $k = 2$ [16].

Let \mathbf{p} be the slippage configuration and let $Q_{\mathbf{p}}$ denote the policy output by the LP. When $k = 2$, the PCS of $Q_{\mathbf{p}}$ for all other vectors $\mathbf{p}' \notin Z_\rho$ is at least ζ . Hence $Q_{\mathbf{p}}$ is a valid policy with respect to the PCS condition. We will prove that $Q_{\mathbf{p}}$ is an optimal policy. For now, we clarify the obstacle that must be overcome to prove optimality. The LP only minimizes the expected number of trials for the vector \mathbf{p} . There is no guarantee that the worst-case expected number of trials of $Q_{\mathbf{p}}$ (outside Z_ρ) occurs at \mathbf{p} . What if the expected number of trials of $Q_{\mathbf{p}}$ on some other vector \mathbf{p}' is much worse? If so, there might be a different policy Q' that costs more than $Q_{\mathbf{p}}$ on \mathbf{p} , but costs less on \mathbf{p}' , and hence has better worst-case cost.

2.2 Definitions and elementary properties

Definition 2.2.1. A *state* $\mathbf{w} = (w_1 \dots, w_k)$ is a nonnegative integer vector denoting that after $\sum_{i=1}^k w_i$ trials alternative i has won w_i times. We write $[\mathbf{w}]$ for the vector containing the entries of \mathbf{w} in nondecreasing order, and $\Pi(\mathbf{w})$ for the set of all distinct permutations of w . For example, $|\Pi(\mathbf{w})| = k!$ iff the entries of $[\mathbf{w}]$ are strictly increasing.

Definition 2.2.2. A *policy* Q is a decision rule which at the end of each trial, chooses whether to terminate or run another trial. A policy must be *neutral*, that is, symmetric with respect to reordering the alternatives, because it has no prior knowledge of the probability vector \mathbf{p} . Upon termination the policy decides which alternative to select.

For a policy Q and a probability vector \mathbf{p} , the *probability of correct selection*, denoted

$PCS^Q(\mathbf{p})$, is the probability that Q selects alternative k for a given \mathbf{p} . Another measure of interest is $E^Q[w; \mathbf{p}]$, the expected number of additional trials Q would run with \mathbf{p} starting from \mathbf{w} , conditioned on Q reaching \mathbf{w} . Thus $E^Q[0; \mathbf{p}]$ is the expected number of trials Q runs with \mathbf{p} .

En route to proving the main result, we reduce the set of policies to be considered to those with certain properties. These entail some additional definitions.

Definition 2.2.3. A policy is *deterministic* if it chooses to terminate deterministically given the outcomes of the trials that have already been run. Otherwise the policy is *randomized*. The policy is *sane* if, upon terminating at state \mathbf{w} , it selects alternative j only if $w_j = [\mathbf{w}]_k$, that is, it never selects an alternative that does not have the most wins. The policies that do not do this are *insane*.

It follows from neutrality that every sane policy, upon termination, selects with equal probability among the alternatives with the largest number of wins.

Definition 2.2.4. A state \mathbf{w} is *terminal* if the probability is strictly positive that Q will terminate at w . A state \mathbf{w} is *reachable* if the probability is strictly positive that Q will reach \mathbf{w} . A state \mathbf{w} is *traversable* if the probability is strictly positive that Q will choose to run a trial at \mathbf{w} . We write T^Q for the set of terminal states of Q .

Definition 2.2.5. A policy is *bounded* if $\exists u$ such that the policy never runs more than u trials. A policy is (*strongly*) *bounded in expectation* if ($\exists u$ such that) $\forall \mathbf{p} \in \mathcal{P}$, $E^Q[0; \mathbf{p}]$ is finite (is at most u).

Definition 2.2.6. A policy is *Markovian* if its termination rule is only based on the current state. Denote the conditional probability that a Markovian policy Q terminates at state \mathbf{w} , given that it arrives at \mathbf{w} , as $Q(\mathbf{w})$.

It follows that every reachable state of a Markovian deterministic policy is either terminal or traversable, but not both. To illustrate the notation, the neutrality of a Markovian policy Q can be expressed as $Q(\mathbf{w}) = Q([\mathbf{w}]) \forall \mathbf{w}$.

Theorem 2.2.1. For all sane policies Q there exists an equivalent sane Markovian randomized policy Q' in the sense that for all probability vectors \mathbf{p} and states \mathbf{w} , P (arriving at \mathbf{w}) and P (terminating at \mathbf{w} |arriving at \mathbf{w}) are the same in both policies. And hence $PCSS^Q(\mathbf{p}) = PCSS^{Q'}(\mathbf{p})$ and $E^Q[0; \mathbf{p}] = E^{Q'}[0; \mathbf{p}]$ for all $\mathbf{p} \in \mathcal{P}$.

Proof. For a non-Markovian deterministic policy Q and $\mathbf{w} \in T^Q$, let $\psi(\mathbf{w})$ be the number of paths (sequences of states) leading to \mathbf{w} that can happen in Q . Let $\phi(\mathbf{w})$ be the number of paths that lead to termination at \mathbf{w} . Let Q' be a policy with $T^{Q'} = T^Q$ and new termination rule as follows: $Q'(\mathbf{w}) = \phi(\mathbf{w}) / \psi(\mathbf{w}) \forall \mathbf{w} \in T^{(Q')}$ independent of other possible terminations.

For a non-Markovian randomized policy Q and $\mathbf{w} \in T^Q$, let $j \in J_{\mathbf{w}}$ index the paths leading to \mathbf{w} , and let $\psi_j(\mathbf{w})$ be the probability that path j reaches \mathbf{w} , i.e., if path j is $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n, \mathbf{w}\}$ then

$$\psi_j(\mathbf{w}) = \prod_{i=1}^n P(\text{not stopping at } \mathbf{w}_i | \text{we reached } \mathbf{w}_i \text{ by taking the path } \mathbf{w}_1, \dots, \mathbf{w}_{i-1}, \mathbf{w}_i)$$

. Let $q_j(\mathbf{w})$ be the probability that Q terminates at \mathbf{w} conditioned on arriving at \mathbf{w} via the path ψ_j . Let Q' be a policy with $T^{Q'} = T^Q$ and new termination rule as follows:

$$Q'(\mathbf{w}) = \frac{\sum_j q_j \psi_j(\mathbf{w})}{\sum_j \psi_j(\mathbf{w})}$$

independent of other possible terminations. □

Therefore, in most of what follows, it will suffice to consider only Markovian policies.

Definition 2.2.7. A randomized policy Q is equivalent to a *mixture* of deterministic policies if there exists a probability measure μ on the set of deterministic policies, Ω , such that Q is equivalent to making an initial choice $Q' \in \Omega$ according to μ and then running the policy Q' .

Theorem 2.2.2. *Every sane policy is equivalent to a mixture of countably many deterministic Markovian sane policies.*

Proof. Denote $|\mathbf{w}| \equiv \sum_{i=1}^k w_i$. By Theorem 2.2.1 there exists a sane Markovian policy Q equivalent to the given policy. Construct a mixture equivalent to Q inductively using a function *mix* as follows. The idea of the construction is to iteratively peel off a fraction of a deterministic policy from Q without adding more randomness to what remains of Q .

Begin procedure mix Let m be the smallest integer such that $|\mathbf{w}| = m$ and $0 < Q(\mathbf{w}) < 1$. Let $q_1 = \min \{Q(\mathbf{w}) : |\mathbf{w}| = m, Q(\mathbf{w}) > 0\}$. Define a deterministic policy Q_1 as follows:

Step 1. Q_1 is identical to Q for all states w with $|\mathbf{w}| < m$.

Step 2. For all \mathbf{w} such that $|\mathbf{w}| = m$, if $p_{\mathbf{w}} \geq q_1$ make \mathbf{w} terminal. Otherwise make \mathbf{w} not terminal.

Step 3. For all $n = m + 1, m + 2, \dots$ inductively classify states \mathbf{w} as terminal or not according to the following rule: For all \mathbf{w} such that $|\mathbf{w}| = n$ and there is a positive probability of being reached with Q_1 given the classification of states $\mathbf{v} : |\mathbf{v}| \leq n - 1$, if $Q(\mathbf{w}) \geq q_1$ make w terminal, otherwise make \mathbf{w} not terminal.

Step 4.

Let \bar{Q}_1 be a randomized policy with

$$\bar{Q}_1(\mathbf{w}) = \begin{cases} \frac{Q(\mathbf{w}) - q_1 Q_1(\mathbf{w})}{1 - q_1} & , \text{ if } Q_1(\mathbf{w}) \text{ is defined} \\ Q(\mathbf{w}) & , \text{ otherwise} \end{cases}$$

End procedure mix

Define $\text{mix}(Q) = \{(Q_1, q_1)\} \cup (1 - q_1) \text{mix}(\bar{Q}_1)$, where $(1 - q_1) \text{mix}(\bar{Q}_1)$ means that we run policies in the $\text{mix}(\bar{Q}_1)$ with probabilities $(1 - q_1) \bar{q}_i$ rather than \bar{q}_i . The new randomized policy \bar{Q}_1 is what remains of Q after peeling off Q_1 . Set $Q = \bar{Q}_1$ and iterate.

At each iteration of this procedure we add at most one deterministic policy per state of

Q with randomized termination rule. Therefore the mixture is countable. Furthermore, by construction this mixture of deterministic policies is equivalent to Q . \square

Corollary 2.2.1. *Every sane randomized policy for which only a finite number of states have a randomized termination rule is equivalent to a mixture of finitely many deterministic Markovian sane policies.*

Proof. Let Q have n states with randomized termination rule, that is, $|\mathbf{w} : 0 < Q(\mathbf{w}) < 1| = n$. Using the construction above, Q is equivalent to a mixture of at most $n + 1$ deterministic policies. \square

The converse to Corollary 2.2.1 is false. For instance, let Q be a policy for two alternatives which stops as soon as each alternative has won at least one trial. Let Q' be a policy which never stops. Run Q with probability $\frac{1}{2}$, otherwise run Q' . In the randomized policy defined by this mixture, for all $k \geq 1$ state $\{1, k\}$ is terminal with termination probability $1 - \frac{1}{k+1}$.

Corollary 2.2.2. *Every bounded sane randomized policy Q is equivalent to a mixture of finitely many bounded sane deterministic Markovian policies.*

Proof. A bounded randomized policy has finitely many states with randomized termination rules and every policy in its deterministic mixture has to be bounded. \square

Theorem 2.2.3. *There exists a sane deterministic policy strongly bounded in expectation which cannot be expressed as a mixture of bounded sane deterministic policies.*

Proof. Let Q be the sane policy for two alternatives which terminates if one of the alternatives has won 2 more trials than the other. By elementary probability we have $E^Q [0; \mathbf{p}] \leq 4$. However, any state of the form (i, i) is reachable and the policy never terminates at (i, i) . Therefore any component of the mixture that occurs with positive probability must be able to run more than $2i$ trials for any i and hence must be unbounded. \square

For any policy Q strongly bounded in expectation, the probability of correct selection PCS^Q gets arbitrarily close to $\frac{1}{n}$ for probability vectors close to $(\frac{1}{k}, \dots, \frac{1}{k})$. As stated previously, those vectors are in the indifference zone Z_ρ , the set of probability vectors such that $\frac{p_k}{p_{k-1}} < \rho$.

Definition 2.2.8. Probability vectors not in Z_ρ are called *permissible*.

The *least favorable configuration (LFC)* for Q is the permissible probability vector \mathbf{p} with the smallest $PCS^Q(\mathbf{p})$.

The *slippage configuration* is defined by $\sum_{i=1}^k p_i = 1, \frac{p_k}{p_i} = \rho, \forall i = 1, \dots, k - 1$. That is, all alternatives except the best one have the same probability, and that probability is the largest possible which makes \mathbf{p} permissible.

[16] proved that, for all sane Markovian policies bounded in expectation with $k = 2$ alternatives, the LFC vector is the slippage configuration. In the remainder of this paper we prove that for two alternatives there is a class of policies such that $E^Q [0; \mathbf{p}]$ is maximized at the slippage configuration for all $\mathbf{p} \notin Z_\rho$. Furthermore, the linear program of [15] run at the slippage configuration produces policies in this class, which therefore minimize the worst case $E^Q [0; \mathbf{p}]$ subject to satisfying all of the input requirements.

2.3 Holey and unholey policies

Everything from here on is restricted to the case of $k = 2$ alternatives. As illustrated in Figure 2.1, Markovian policies for $k = 2$ can be readily visualized in 2 dimensions. The two elements (w_1, w_2) of state w define the horizontal and vertical components. For example, neutrality is equivalent to reflective symmetry across the diagonal $w_1 = w_2$.

Definition 2.3.1. A deterministic policy is *not curtailed* if it has a traversable state \mathbf{w} whose terminal descendants all make the same selection or are equally likely to choose either of the alternatives. A mixture of deterministic policies is not curtailed if one or more of its components is not curtailed.

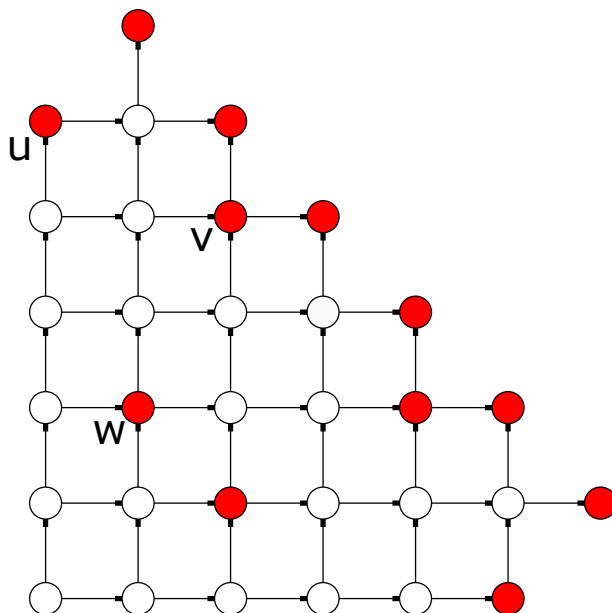


Figure 2.1: **Example of a policy.** All red states are terminal. All white states are traversable. Unreachable states are not shown. u and v are terminal but not holes as they have untraversable successors in at least one direction. w is a hole as both $w + (1, 0)$ and $w + (0, 1)$ are traversable.

(Definition 2.3.1 is worded ambiguously as to whether or not either all descendants must make the same selection or all must choose randomly. By the symmetry of neutrality the two choices are equivalent.)

Definition 2.3.2. A terminal state $w = (w_1, w_2)$ in a deterministic policy is a *hole* if there exist strictly positive integers m_1, m_2 such that $\mu_1 = (w_1 + m_1, w_2)$ and $\mu_2 = (w_1, w_2 + m_2)$ are traversable. A policy with a hole is *holey*; a policy without holes is *unholey*.

It follows from neutrality that if a terminal (w_1, w_2) with $w_1 \leq w_2$ is not a hole, then either $(w_1, w_2 + n)$ and $(w_2 + n, w_1)$ are not traversable for all $n \geq 1$, or $(w_1 + n, w_2)$ and $(w_2, w_1 + n)$ are not traversable for all $n \geq 1$, or both. For example, see Figure 2.1.

A policy that is not curtailed is obviously suboptimal with respect to $E^Q[0, \mathbf{p}]$ for all \mathbf{p} . Moreover, it is obvious how to alter such a policy so that it is better for all \mathbf{p} . Holes turn out to be more complicated. Intuitively a holey policy should be suboptimal. Consider

for example the policy that terminates if an alternative has four wins but also terminates at $\mathbf{w} : [\mathbf{w}] = (1, 2)$ (See Figure 2.2). It seems stupid not to run a test at $(1, 2)$ but to run one at $(1, 3)$ because the PCS at the latter is larger. Additional information at the former should therefore be more valuable than at the latter. Unfortunately, the simple alteration that makes $(1, 2)$ traversable and $(1, 3)$ terminal can fail to work. That alteration changes both the PCS and $E^Q[\mathbf{0}, \mathbf{p}]$. Worse, the changes vary with \mathbf{p} because \mathbf{p} affect the probability of reaching these states. The following conditions for suboptimality are therefore limited, and the corresponding alteration depends on \mathbf{p} .

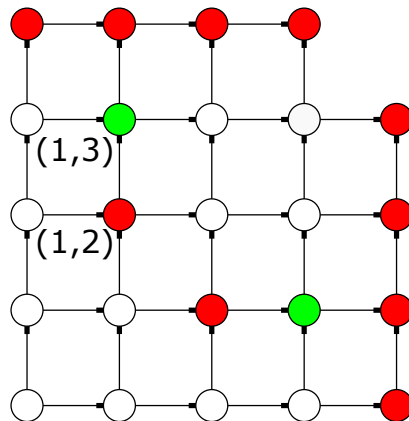


Figure 2.2: **Policy that terminates if an alternative has four wins but also terminates at $\mathbf{w} : [\mathbf{w}] = (1, 2)$.**

Theorem 2.3.1. *Suppose a deterministic, sane, and curtailed Markovian policy Q is holey. Suppose further that Q has finite number of terminal states and each terminal state \mathbf{w} has finite number of non-terminal states on its positive diagonal*

$$D_{\mathbf{w}} := \{\mathbf{u} : \mathbf{u} = \mathbf{w} + n(1, 1), n \geq 1\}.$$

Then for any probability vector $\mathbf{p} \in \mathcal{P}$ there exists a Markovian sane curtailed policy with the same PCS but strictly lower expected number of trials at \mathbf{p} .

Proof. The proof is by induction on the maximum number of non-terminal states on the diagonals of holes.

Suppose that for each hole \mathbf{w} in Q , at least one of the states on $D_{\mathbf{w}}$ is terminal and the rest are either terminal or unreachable. By definition of a hole, there exists m such that $\mathbf{v} = [\mathbf{w}] + (0, m)$ is non-terminal. However none of the paths from \mathbf{v} can cross the diagonal $D_{[\mathbf{w}]}$ and hence the policy never changes its decision if run from \mathbf{v} , making the policy not curtailed.

Suppose the result holds for all policies with at most at most n non-terminal states on the diagonals $D_{[\mathbf{w}]}$ of each hole $[\mathbf{w}]$. Let Q be a policy with at most $n + 1$ non-terminal states on the diagonals of its holes. Let $[\mathbf{w}_1], [\mathbf{w}_2], \dots, [\mathbf{w}_s]$ be the holes that have exactly $n + 1$ non-terminal states on their diagonals.

Let r_i be positive integers such that $[\mathbf{v}_i] := [\mathbf{w}_i] + r_i(1, 1)$ are non-terminal for all $i \in \{1, 2, \dots, s\}$. Let $P(\mathbf{w})$ be the probability of correct selection conditioned on being at a state in $\Pi(\mathbf{w})$, and $\psi(\mathbf{w})$ be the probability of reaching $\Pi(\mathbf{w})$ for the given \mathbf{p} . By a small abuse of notation, let $E(\Pi(\mathbf{w}))$ be the expected number of number of additional steps until termination conditioned on being at $\Pi(\mathbf{w})$ for the given \mathbf{p} by running a sub-policy $Q_{\mathbf{w}}$.

Define a non-Markovian randomized policy R as follows: make states $\Pi(\mathbf{v}_i)$ terminal with probability ρ_i ; independently with probability α_i make $\Pi(\mathbf{w}_i)$ non-terminal and run $Q_{[\mathbf{v}_i]}$'s from them; otherwise the policy is the same as Q . By Theorem 2.2.1 there exists a Markovian policy R' equivalent to R . We have that:

$$\begin{aligned}
PCS^Q(\mathbf{p}) - PCS^{Q'}(\mathbf{p}) &= \sum_{i=1}^s \alpha_i \psi(\mathbf{w}_i) (P(\mathbf{w}_i) - P(\mathbf{v}_i)) \\
&\quad + \sum_{i=1}^s \rho_i \psi(\mathbf{v}_i) (P(\mathbf{v}_i) - P(\mathbf{w}_i)) \\
E^Q[\mathbf{0}; \mathbf{p}] - E^{Q'}[\mathbf{0}; \mathbf{p}] &= - \sum_{i=1}^s \alpha_i \psi(\mathbf{v}_i) E(\Pi(\mathbf{v}_i)) \\
&\quad + \sum_{i=1}^s \rho_i \psi(\mathbf{w}_i) E(\Pi(\mathbf{v}_i))
\end{aligned}$$

For each i there are two possibilities: $\psi(\mathbf{v}_i) / \psi(\mathbf{w}_i) \leq 1$ or $\psi(\mathbf{v}_i) / \psi(\mathbf{w}_i) > 1$. In the

first case set $\alpha_i = \psi(\mathbf{v}_i) / \psi(\mathbf{w}_i)$, $\rho_i = 1$; otherwise $\alpha_i = 1$, $\rho_i = \psi(\mathbf{w}_i) / \psi(\mathbf{v}_i)$. This way we get $PCS^Q(\mathbf{p}) = PCS^{Q'}(\mathbf{p})$ and $E^Q[\mathbf{0}; \mathbf{p}] = E^{Q'}[\mathbf{0}; \mathbf{p}]$

There must exist a component in the mixture of R' , call it \bar{R} , which happens with positive probability and corresponds to the following: For each i such that $\psi(\mathbf{v}_i) / \psi(\mathbf{w}_i) \leq 1$, $\Pi(\mathbf{w}_i)$ are holes and have at least one less non-terminal state along their diagonals, namely $\Pi(\mathbf{v}_i)$. For each i such that $\psi(\mathbf{v}_i) / \psi(\mathbf{w}_i) > 1$, $\Pi(\mathbf{w}_i)$ are non-terminal but $\Pi(\mathbf{v}_i)$ are terminal. The latter, if they are holes, have strictly fewer non-terminal diagonal successors than states in $\Pi(\mathbf{w}_i)$. The rest of the states \mathbf{w} correspond to the event that possible terminations introduced by the shifted sub-policies $Q_{\mathbf{v}_i}$ did not happen. Therefore \bar{R} only has holes with at most n diagonal successors and by the inductive hypothesis there exists a policy \hat{R} which has the same PCS as \bar{R} but strictly lower expected number of trials at \mathbf{p} . Let Q' be a policy which is the same as R' except for running \hat{R} instead of \bar{R} . This happens with positive probability and hence Q' has the same PCS as Q but a strictly smaller expected number of trials. Finally, by construction Q' is curtailed. \square

One can generalize a hole in a natural way for $k \geq 3$ alternatives by looking at non-terminal states along any two directions. However, the proposition doesn't work with this definition. For $k \geq 3$ even if all $\mathbf{w} + n\mathbf{1}$, $n \geq 0$ are terminal, the policy is not prevented from reaching states that make a different selection.

Corollary 2.3.1. *If a deterministic, bounded, sane and curtailed Markovian policy Q is holey then for any given probability vector $\mathbf{p} \in \mathcal{P}$ there exists a bounded sane curtailed Markovian policy with the same PCS but strictly lesser expected number of trials at \mathbf{p} .*

Proof. Bounded policies have finite number of non-terminal states along the diagonals of their holes. \square

Lemma 2.3.1. *Let a Markovian bounded deterministic policy Q be at least one of the following: insane, holey or not curtailed. Then for all probability vectors $\mathbf{p} \notin Z_\rho$, there exists a bounded, sane, unholey and curtailed policy Q' such that,*

$$\begin{aligned}
PCS^Q(\mathbf{p}) &\leq PCS^{Q'}(\mathbf{p}) \\
E^Q[0; \mathbf{p}] &\geq E^{Q'}[0; \mathbf{p}]
\end{aligned}$$

and at least one of these inequalities holds strictly.

Proof. If the policy is insane, we can keep the same expected number of trials but achieve better strictly better PCS by making it sane.

If the policy is not curtailed then we can keep the PCS constant and strictly lower the expected number of trials by terminating at the non-terminal states whose terminal descendants all make the same selection.

If the policy is holey then we can strictly lower the expected number of trials by Theorem 2.3.1. □

We remark that Lemma 2.3.1 is false if Q' is required to be deterministic. A counterexample is given in the appendix (Section B).

Theorem 2.3.2. *Let a Markovian bounded deterministic policy Q be sane, curtailed and unholey. Let \mathbf{w} be a terminal state with maximal $|\mathbf{w}|$. Then $|\mathbf{w}|$ is odd.*

Proof. Suppose a policy Q runs at most n trials. Let $[\mathbf{w}] = (w_1, w_2)$ be a terminal state such that $||[\mathbf{w}]|| = n$. Since $[\mathbf{w}]$ is terminal, at least one of its two predecessor states $(w_1 - 1, w_2), (w_1, w_2 - 1)$ must be traversable. If $w_2 \geq w_1 + 2$ then by maximality of $|\mathbf{w}|$ both successors of both predecessor make the same decision as $[\mathbf{w}]$ and hence the policy is not curtailed. If $w_1 = w_2$, neutrality and maximality of $|\mathbf{w}|$ imply that making both predecessors terminal would not change the decision probability. Therefore, $w_2 = w_1 + 1$ and the policy runs at most an odd number of trials. □

Theorem 2.3.3. *Let a bounded Markovian deterministic policy Q be sane, curtailed and unholey. Then for any pair of distinct terminal states $[\mathbf{w}], [\mathbf{v}] \in T^Q$, $w_1 \neq v_1$ and $w_1 < v_1 \Rightarrow w_2 \leq v_2$. Hence these inequalities impose a strict total order on all $[\mathbf{w}] \in T^Q$.*

Proof. Let $\mathbf{w} = [\mathbf{w}]$ and $\mathbf{v} = [\mathbf{v}]$. State \mathbf{v} is not a hole because Q is unholey. Hence (i) $\mathbf{v} + (0, n)$ is not traversable $\forall n \geq 0$, or (ii) $(v_2, v_1) + (0, n)$ is not traversable $\forall n \geq 0$.

If $w_1 = v_1$ then by symmetry assume $w_2 > v_2$. Since \mathbf{w} is reachable it has a traversable immediate predecessor \mathbf{u} , possibly $\mathbf{u} = (w_1 - 1, w_2)$ or $\mathbf{u} = (w_1, w_2 - 1)$ (or both). We will show that both possibilities contradict curtailment.

Case $\mathbf{u} = (w_1 - 1, w_2)$ traversable: If (i) then by $u_1 < v_1, u_2 \geq v_2$, policy Q cannot cross the main diagonal $(n, n) : n \geq 0$ starting from \mathbf{u} , and so \mathbf{u} being traversable contradicts Q being curtailed and sane. (Geometrically, \mathbf{u} is to the left and at least as high as \mathbf{v} , so Q can't cross the upward ray from \mathbf{v} . By sanity, Q must cross the main diagonal to select alternative 1 instead of 2.) Similarly, if (ii) then by $u_1 < v_2, u_2 \geq v_1$, policy Q cannot cross the main diagonal from \mathbf{u} , contradicting curtailment or sanity.

Case $\mathbf{u} = (w_1, w_2 - 1)$ traversable: If (i) there is an immediate contradiction because $u_1 = w_1 = v_1$ and $u_2 = w_2 - 1 \geq v_2$. If (ii) we have $u_1 = w_1 = v_1 \leq v_2$ and $u_2 = w_2 - 1 \geq v_2 \geq v_1$. As in the previous case, policy Q starting at \mathbf{u} cannot traverse $(v_2, v_1) + (0, n) : n \geq 0$ to get past the main diagonal, so by curtailment and sanity of Q we have a contradiction with the traversability of \mathbf{u} .

So far we have proved $w_1 \neq v_1$. For the rest of the proof, suppose $w_1 < v_1$ but $w_2 > v_2$ as depicted in Figure 2.3. Again, at least one choice of $\mathbf{u} = (w_1 - 1, w_2)$ or $\mathbf{u} = (w_1, w_2 - 1)$ is traversable. However, neither choice of \mathbf{u} can cross either of the rays defined by (i) or (ii), a contradiction. \square

Corollary 2.3.2. *Let a bounded deterministic Markovian policy Q be sane, curtailed and unholey. Let $[\mathbf{w}] = (w_1, w_2)$, $w_1 \neq w_2$ be traversable. Then $[\mathbf{w}] + (1, 0)$ is traversable.*

Proof. By hypothesis, state $[\mathbf{w}] + (1, 0)$ is reachable from traversable state $[\mathbf{w}]$. Hence

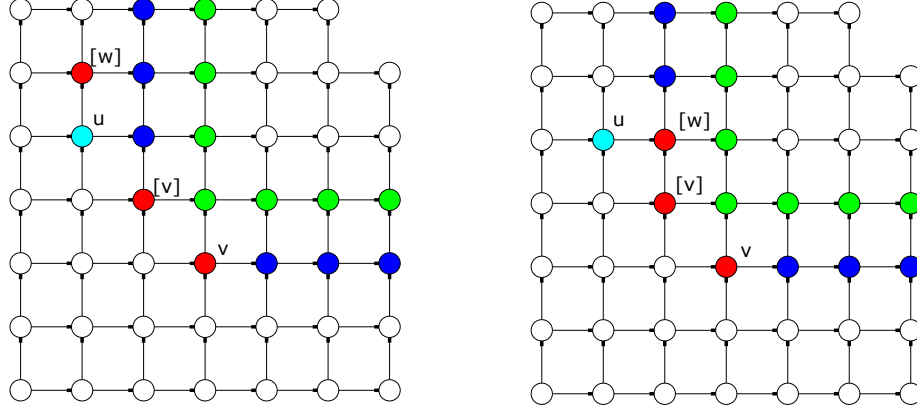


Figure 2.3: **Two possibilities for terminal states to block u from changing its decision.** At least one of the green or blue sets of dots must be terminal as $[v]$ is not a hole. On the left $w_1 < v_1$ and $w_2 > v_2$. On the right $w_1 = v_1$ and $v_2 < w_2$

it suffices to prove that state $[w] + (1, 0)$ is not terminal. Suppose to the contrary it is terminal. Then by Theorem 2.3.3 no state $[w] + n(0, 1) : n \geq 1$ can be terminal. Since $[w]$ is traversable, by induction all states $[w] + n(0, 1) : n \geq 1$ are traversable. Then Q is not bounded, a contradiction. \square

A consequence of Theorem 2.3.3 and Corollary 2.3.2 is that the set of terminal states of bounded, sane, unholey and curtailed policies form monotonic step functions and the policies are defined by them and the bounded region between them. This geometric property is a key step towards our main result, and might be generalizable to higher dimensions.

2.4 Slipping to optimality

We will prove optimality at the policy for the slippage configuration \mathbf{p}_{SC} by gradually changing a policy for an arbitrary vector $\mathbf{p} \notin Z_\rho$ to one for \mathbf{p}_{SC} while monotonically changing the expected number of trials. Whenever \mathbf{p} is replaced by \mathbf{p}_{SC} , the policy becomes more likely to move towards the main diagonal from states (w_1, w_2) with $w_1 < w_2$ (above the main diagonal), and more likely to move away from the main diagonal from states with $w_1 > w_2$ (below the main diagonal) because $\mathbf{p}_{SC_1} > \mathbf{p}_1$. The proof would be easier if the policy were always more likely to move towards the main diagonal, because

more trials are needed to achieve a desired PCS from states near the diagonal. Instead, changing the policy has an asymmetric effect about the main diagonal. Fortunately, as the next lemmas show, the benefit above the diagonal outweighs the disbenefit below the diagonal because $\mathbf{p}_{SC_1} < \mathbf{p}_{SC_2}$. Then, protected by the geometric property from Theorem 2.3.3, we find how to sequence the gradual changes from \mathbf{p} to the slippage vector \mathbf{p}_{SC} and maintain monotonicity.

Lemma 2.4.1. *Let a bounded deterministic Markovian policy Q be sane, curtailed and unholey and \mathbf{p} be a probability vector outside the indifference zone. As before let $E^Q[\mathbf{w}, \mathbf{p}]$ be the expected number of additional steps taken by policy Q until termination, conditioned on being at a state \mathbf{w} for the vector \mathbf{p} . Then for any $\mathbf{w}_1^+ = (w, w + k + 1)$, $\mathbf{w}_1^- = (w + k + 1, w)$, $\mathbf{w}_2^+ = (w + k, w + 1)$, $\mathbf{w}_2^- = (w + 1, w + k)$ where w, k are non-negative integers $\mathbf{w}_2^+, \mathbf{w}_2^-$ non-terminal in Q , the following inequalities hold:*

$$p^k E^Q[\mathbf{w}_2^-; \mathbf{p}] \geq (1 - p)^k E^Q[\mathbf{w}_2^+; \mathbf{p}] \quad (2.1)$$

$$p^{k+2} E^Q[\mathbf{w}_1^-; \mathbf{p}] \geq (1 - p)^{k+2} E^Q[\mathbf{w}_1^+; \mathbf{p}] \quad (2.2)$$

$$p^k (E^Q[\mathbf{w}_2^-; \mathbf{p}] - E^Q[\mathbf{w}_1^-; \mathbf{p}]) \geq (1 - p)^k (E^Q[\mathbf{w}_2^+; \mathbf{p}] - E^Q[\mathbf{w}_1^+; \mathbf{p}]) \quad (2.3)$$

Proof. The proof is by induction on m , where $|\mathbf{w}_2^+| = \max\{|\mathbf{v}| : \mathbf{v} \in T^Q\} - m$.

The base case $m = 1$ is immediate. Suppose the result holds for $m - 1$. Let $\mathbf{v}_1^+ = (w, w + k + 2)$, $\mathbf{v}_2^+ = (w + 1, w + k + 1)$, $\mathbf{v}_3^+ = (w + 2, w + k)$ and $\mathbf{v}_1^- = (w + k + 2, w)$, $\mathbf{v}_2^- = (w + k + 1, w + 1)$, $\mathbf{v}_3^- = (w + k, w + 2)$. See Figure 2.4. There are two cases, depending on whether $\mathbf{w}_1^+, \mathbf{w}_1^-$ are terminal or not.

In the first case, inequality 2.2 holds trivially. The inequalities 2.1 and 2.3 are equivalent

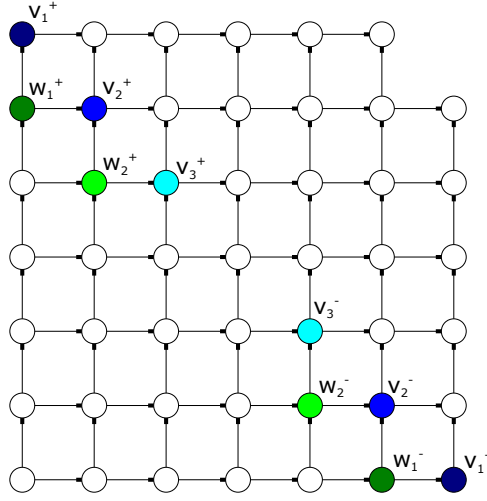


Figure 2.4: **The inductive step in Lemma 2.4.1**

and we get:

$$\begin{aligned}
& p^k E^Q [\mathbf{w}_2^-, \mathbf{p}] \\
& - (1-p)^k E^Q [\mathbf{w}_2^+, \mathbf{p}] = p^k (1 + p E^Q [\mathbf{v}_2^-, \mathbf{p}] + (1-p) E^Q [\mathbf{v}_3^-, \mathbf{p}]) \\
& \quad - (1-p)^k (1 + p E^Q [\mathbf{v}_2^+, \mathbf{p}+] + (1-p) E^Q [\mathbf{v}_3^+, \mathbf{p}]) \\
& = (p^k - (1-p)^k) + (p^{k+1} E^Q [\mathbf{v}_2^-, \mathbf{p}] - (1-p)^{k+1} E^Q [\mathbf{v}_2^+, \mathbf{p}]) \\
& \quad + p(1-p) (p^{k-1} E^Q [\mathbf{v}_3^-, \mathbf{p}] - (1-p)^{k-1} E^Q [\mathbf{v}_3^+, \mathbf{p}]) \\
& \geq 0
\end{aligned}$$

The first term is non-negative as $p > (1-p)$. The other terms are non-negative by the inductive hypothesis applied to \mathbf{v}_3^+ and \mathbf{v}_3^- by Corollary 2.3.2.

If $\mathbf{w}_1^+, \mathbf{w}_1^-$ are non-terminal, then the proof of inequality 2.1 is the same as above and

2.2 is analogous. For 2.3 we have:

$$\begin{aligned}
& p^k (E^Q [\mathbf{w}_2^-, \mathbf{p}] - E^Q [\mathbf{w}_1^-, \mathbf{p}]) \\
& + (1-p)^k (E^Q [\mathbf{w}_1^+, \mathbf{p}] - E^Q [\mathbf{w}_2^+, \mathbf{p}]) = p^{k+1} E^Q [\mathbf{v}_2^-, \mathbf{p}] + p^k (1-p) E^Q [\mathbf{v}_3^-, \mathbf{p}] \\
& \quad - p^{k+1} E^Q [\mathbf{v}_1^-, \mathbf{p}] - p^k (1-p) E^Q [\mathbf{v}_2^-, \mathbf{p}] \\
& \quad + (1-p)^{k+1} E^Q [\mathbf{v}_1^+, \mathbf{p}] + (1-p)^k p E^Q [\mathbf{v}_2^+, \mathbf{p}] \\
& \quad - (1-p)^{k+1} E^Q [\mathbf{v}_2^+, \mathbf{p}] - (1-p)^k p E^Q [\mathbf{v}_3^+, \mathbf{p}] \\
& = p^{k+1} (E^Q [\mathbf{v}_2^-, \mathbf{p}] - E^Q [\mathbf{v}_1^-, \mathbf{p}]) \\
& \quad + (1-p)^{k+1} (E^Q [\mathbf{v}_1^+, \mathbf{p}] - E^Q [\mathbf{v}_2^+, \mathbf{p}]) \\
& \quad + p(1-p)p^{k-1} (E^Q [\mathbf{v}_3^-, \mathbf{p}] - E^Q [\mathbf{v}_2^-, \mathbf{p}]) \\
& \quad + p(1-p)(1-p)^{k-1} (E^Q [\mathbf{v}_2^+, \mathbf{p}] - E^Q [\mathbf{v}_3^+, \mathbf{p}]) \\
& \geq 0
\end{aligned}$$

by inductive hypothesis applied to \mathbf{v}_2^+ , \mathbf{v}_3^+ , \mathbf{v}_2^- , and \mathbf{v}_3^- which are non-terminal by Corollary 2.3.2. □

Lemma 2.4.2. *Let a bounded deterministic Markovian policy Q be sane, curtailed and unholey and \mathbf{p} be a probability vector outside the indifference zone. Then for all non-terminal $\mathbf{w} = (w_1, w_2)$ with $w_1 \geq w_2$ we have $E[\mathbf{w}', \mathbf{p}] \leq E^Q(\mathbf{w}, \mathbf{p})$, where $\mathbf{w}' = \mathbf{w} + (1, -1)$.*

Proof. The proof is again by induction on m , where $|\mathbf{w}| = \max\{|\mathbf{v}| : v \in T^Q\} - m$. If \mathbf{w}' is terminal, the result holds trivially. Furthermore, if $w_1 = w_2 + 1$ then the result holds by Lemma 2.4.1. Otherwise, let $\mathbf{v}_3 = \mathbf{w} + (0, 1)$, $\mathbf{v}_2 = \mathbf{w} + (1, 0) = \mathbf{w}' + (0, 1)$ and $\mathbf{v}_1 = \mathbf{w}' + (1, 0)$. Then for all \mathbf{v}_i , $(\mathbf{v}_i)_1 \geq (\mathbf{v}_i)_2$ and hence, by the inductive hypothesis

$$E(\mathbf{v}_3) \geq E(\mathbf{v}_2) \geq E(\mathbf{v}_1)$$

For \mathbf{w} and \mathbf{w}' , looking at the next step we get:

$$\begin{aligned}
E[\mathbf{w}, \mathbf{p}] - E[\mathbf{w}', \mathbf{p}] &= (1 + (1 - p) E[\mathbf{v}_3, \mathbf{p}] + pE[\mathbf{v}_2, \mathbf{p}]) \\
&\quad - (1 + (1 - p) E[\mathbf{v}_2, \mathbf{p}] + pE[\mathbf{v}_1, \mathbf{p}]) \\
&= (1 - p) (E[\mathbf{v}_3, \mathbf{p}] - E[\mathbf{v}_2, \mathbf{p}]) + p(E[\mathbf{v}_2, \mathbf{p}] - E[\mathbf{v}_1, \mathbf{p}]) \\
&\geq 0
\end{aligned}$$

□

Lemma 2.4.3. *Let a bounded Markovian deterministic policy Q be sane, curtailed and unholey. Let \mathbf{p} be a probability vector outside the indifference zone. Let \mathbf{s} be the slippage probability vector. Let S be a set of 'slippery' states in Q such that if $\mathbf{w} \in S$ then $\Pi(\mathbf{w}) \subseteq S$ and for all integers $i > 0$ and $j > 0$, $\mathbf{w} + (i, j) \in S$. At a slippery state, we use \mathbf{s} instead of \mathbf{p} . Suppose Q' is a sane policy with $T^{Q'} = T^Q$ and suppose \mathbf{w} is a state such that all of its successors are slippery in Q . Suppose the set of slippery states in Q' is $S \cup \Pi(\mathbf{w})$. Then $E^Q[0; \mathbf{p}]$, the expected number of trials run by Q , does not exceed $E^{Q'}[0; \mathbf{p}]$, the expected number of trials run by Q' .*

Proof. Proof If \mathbf{w} is terminal, the result is immediate. Otherwise, let $\psi(\mathbf{w})$ be the probability of reaching states in $\Pi(\mathbf{w})$ (it is the same in Q and Q'). Let $k = |\mathbf{w}_2 - \mathbf{w}_1|$. Let $\mathbf{v}_1^+ = [\mathbf{w}] + (0, 1)$, $\mathbf{v}_1^- = [\mathbf{w}] + (1, 0)$, $\mathbf{v}_2^+ = [\mathbf{w}] + (1, 0)$, $\mathbf{v}_2^- = [\mathbf{w}] + (0, 1)$. Then the difference in expectations is:

$$\begin{aligned}
E^{Q'}[0; \mathbf{p}] - E^Q[0; \mathbf{p}] &= \psi(\Pi(\mathbf{w})) \frac{p - s}{p^k + (1 - p)^k} p^k (E^Q[\mathbf{v}_2^-; \mathbf{p}] - E^Q[\mathbf{v}_1^-; \mathbf{p}]) \\
&\quad + \psi(\Pi(\mathbf{w})) \frac{p - s}{p^k + (1 - p)^k} (1 - p)^k (E^Q[\mathbf{v}_1^+; \mathbf{p}] - E^Q[\mathbf{v}_2^+; \mathbf{p}])
\end{aligned}$$

By Lemma 2.4.1 we know that

$$s^k (E^Q[\mathbf{v}_2^-; \mathbf{p}] - E^Q[\mathbf{v}_1^-; \mathbf{p}]) \geq (1 - s)^k (E^Q[\mathbf{v}_2^+; \mathbf{p}] - E^Q[\mathbf{v}_1^+; \mathbf{p}])$$

□

Furthermore by Lemma 2.4.2, $E^Q [\mathbf{v}_2^-, \mathbf{p}] - E^Q [\mathbf{v}_1^-, \mathbf{p}] \geq 0$. If the RHS is negative then changing s to p will keep the inequality valid. Otherwise, we have that $p > s$, so changing s to p will increase the LHS and decrease the RHS and, as they are both positive, the inequality will also be preserved. Therefore:

$$E^{Q'} [\mathbf{0}; \mathbf{p}] \geq E^Q [\mathbf{0}; \mathbf{p}]$$

Theorem 2.4.1. *Let a bounded deterministic policy Q be sane, curtailed and unholey. Then over all $\mathbf{p} = (1 - p, p)$ outside the indifference zone, $E^Q [\mathbf{0}; \mathbf{p}]$ is maximized at the slippage configuration.*

Proof. Start with Q and an empty set of slippery states $S = \emptyset$. Then repeatedly add a state w to S with $|w|$ maximal among states not yet in S , which addition by Lemma 2.4.3 does not decrease the expected number of trials. Since Q is bounded, after a finite number of additions we end up with a policy Q' which uses the slippage configuration vector at all states with

$$E^{Q'} [\mathbf{0}; \mathbf{p}] = E^Q [\mathbf{0}; \mathbf{s}] \geq E^Q [\mathbf{0}; \mathbf{p}]$$

□

These conditions (with the hole generalization discussed above) are not sufficient for this theorem for $k \geq 3$ alternatives. For instance, any policy which terminates as soon as more than two alternatives have each won at least one trial but is otherwise reasonable will *minimize* its expected number of trials at the slippage configuration.

Theorem 2.4.2. *For two alternatives let \bar{Q} be the policy defined by an optimal solution to the LP of [15] with the following parameters: the minimum required PCS value ζ , the indifference zone parameter ρ , the upper bound u on number of trials, and the slippage configuration \mathbf{p}_{SC} for ρ . Then \bar{Q} has the minimum possible worst-case expected number of*

trials, where the minimum is taken over all policies with upper bound u and PCS at least ζ , and the worst case is taken over all probability vectors outside the indifference zone defined by ρ .

Proof. By Theorem 2.2.1 it is sufficient to prove that \bar{Q} is optimal among the set of Markovian policies. By the correctness of the formulation [15] we have that \bar{Q} is sane and runs at most u trials. Together with [16] it also has the required PCS guarantee. By Corollary 2.2.2, \bar{Q} is a mixture of finitely many sane Markovian deterministic policies. From Lemma 2.3.1 we have that every policy in the mixture of \bar{Q} is curtailed and unholley. By Theorem 2.4.1,

$$\max_{\mathbf{p} \notin Z_\rho} E^{\bar{Q}}[\mathbf{0}; \mathbf{p}] = E^{\bar{Q}}[\mathbf{0}; \mathbf{p}_{SC}],$$

as the expectation of every policy in its mixture is maximized at the slippage configuration. Therefore, any policy Q satisfying the upper bound on the number of trials u and the PCS requirement for the given ρ we have:

$$E^{\bar{Q}}[\mathbf{0}; \mathbf{p}_{SC}] \leq E^Q[\mathbf{0}; \mathbf{p}_{SC}] \leq \max_{\mathbf{p} \notin Z_\rho} E^Q[\mathbf{0}; \mathbf{p}]$$

The first inequality follows from the optimality of \bar{Q} for the LP with the vector \mathbf{p}_{SC} over all policies Q with bounds u and ζ .

□

2.5 Conclusions

The resolution of the multinomial selection problem for $k = 2$ alternatives given here illustrates the usefulness of game-theoretic concepts and linear optimization in a different field of study. Extending the result to $k \geq 3$ alternatives would seem to require both a different conception of holeyness, and a more general theorem than in [16] on the slippage configuration. Although it is straightforward to verify that the slippage configuration satis-

fies the first-order Kuhn-Tucker conditions, achieving the latter does not appear to be easy. At the least it requires the additional condition that the policy be invariant to permuting the alternatives, e.g., that it not know to terminate as soon as the best alternative has the most wins.

Another natural open question is whether or not the integer programming model corresponding to the LP [15] always finds an optimal deterministic policy for $k = 2$. By enumeration and brute-force grid search we have verified an affirmative answer to this question for budget limits up to 12. However, the counterexample to Lemma 2.3.1 in Section B shows that the proof technique employed here cannot apply.

Part II

Epidemiology

CHAPTER 3

BASIC REPRODUCTION NUMBER: THEORETICAL ANALYSIS

3.1 Introduction

Mathematical models of contagion have been used to forecast the likely pattern of disease spread during an outbreak and to inform policy [49, 50, 51]. Classical models divide the population into compartments based on their role in disease transmission, *e.g.*, susceptible, contagious but asymptomatic, contagious and symptomatic, removed. In the classical models, contacts between individuals are spatially homogeneous in the sense that an individual is equally likely to interact with any other individual in the population.

For these models, the reproduction number (R_0) is the standard summary measure of pandemic risk [52]. R_0 is essentially the expected number of people who would be infected directly by a newly infectious person in an otherwise completely susceptible population. R_0 has the crucial property in these models that if $R_0 < 1$, the disease will die out on its own. If $R_0 > 1$ the disease will spread to epidemic levels within the population. It has been used by health organizations both as a summary statistic for the potential spread of the epidemic and to inform public health interventions [25, 26, 27, 53, 54]. There have been some criticisms of R_0 in spatially homogeneous models [55, 56] regarding the difficulty of its computation and its amenability to precise definition. Nonetheless, R_0 remains the standard measure for these models.

More recent models incorporate networks of contacts to account for the non-homogeneous structure of real social interactions. For example, an individual is quite likely to be in physical proximity to several family members, co-workers, and friends, yet quite unlikely to interact with 99% of the people living within a 1-mile radius. The reproduction number has been employed as a summary measure for such models. In general, it is calculated

efficiently by directly counting secondary cases generated by a patient zero [57, 58, 59, 60, 61].

However, the reproduction number lacks both theoretical and empirical justification for contact network models, and several concerns about its usefulness are reported in the literature. It does not seem to have a clear relationship to the final size of the outbreak in real networks [28, 62]. Nor is it clear that R_0 provides a threshold for epidemic-level contagion. It has been observed that it might not be a sufficient condition for a large outbreak [63, 64]. Thus its predictive power is questionable. R_0 has also been criticized as being ambiguous and difficult to compute, there being no prototypical patient zero in a non-homogeneous model [65, 24, 66, 67, 68].

In this chapter, we substantiate these concerns by demonstrating grave theoretical deficiencies of the R_0 measure for network models, and we propose a practicable alternative measure.

Our theoretical analysis explicates the mathematical reasons why the reproduction number is problematic for network models. When disease propagation is not homogeneous, R_0 loses fundamental properties that give it predictive power in homogeneous models. We prove that R_0 does not have a threshold property in either direction. That is, $R_0 > 1$ is neither a sufficient nor necessary condition for an epidemic outbreak. Moreover, the magnitude of R_0 can have an arbitrarily unreliable relationship to the size of the outbreak. Fortunately, there is a good alternative to the reproduction number. Although the exact expected value of the fraction of population that will be infected is computationally intractable, we prove it can be efficiently approximated with high probability to any specified level of accuracy.

3.2 Theoretical Analysis

For clarity, we present a simple epidemic model on a graph. The exposition here extends straightforwardly to more general models.

Each member of the population is represented by a node $v \in V$ of a directed graph $G = (V, E)$. An edge $(v, u) \in E$ in the graph G represents the possibility that v may expose u to the disease. An epidemic is modeled as a stochastic process that evolves on G in discrete time steps indexed by $t = 0, 1, 2, \dots$. At each time step t , each node is in one of three states: Susceptible (S), Infectious (I), and Removed (R). Initially, during step $t = 0$, one or several nodes are infectious, and the rest are susceptible. Every node v that is infectious at step t causes its neighboring susceptible nodes to become infectious at step $t+1$ with probability p . To be precise, if v is infectious during step t , u is susceptible during step t , and $(v, u) \in E$, then with probability p , independent of other edges in E , v causes u to be infectious in step $t + 1$. By independence, if a susceptible node u is connected by edges from k infectious nodes at time t , then u will become infectious at time $t + 1$ with probability $1 - (1 - p)^k$. In this simple model, nodes stay infected for exactly one time step, and then become removed. Our case study in the subsequent chapter employs a more complex model, in which p varies depending on the edge and the time spent in the infectious state, which itself is also variable.

Suppose the graph starts with one infectious node s at time $t = 0$. Let $S(t)$, $I(t)$ and $R(t)$ be respectively, the number of susceptible, infectious and removed nodes at time t . Write $R(\infty)$ for the number of removed nodes after the process ends. The basic reproduction number is defined as the expected number of secondary cases produced by a single typical infection in a completely susceptible population [52]. For this model, assuming that the initial infected node is chosen uniformly at random, the basic reproduction number takes a simple form:

$$R_0 = \sum_{v \in V} \frac{\delta(v) p}{|V|} = \frac{|E| p}{|V|},$$

where $\delta(v)$ is the out-degree of a vertex.

We now state computational properties of $R(\infty)$. In some ODE models (in particular, the one obtained from the spatially homogeneous model, e.g. Reed-Frost under an appropriate fluid limit), one can directly compute the expected final epidemic size from a

given R_0 [52]. However, for non-homogeneous network models, $R(\infty)$ has a more severe level of computational complexity than the notorious NP-complete problems. Computing $E[R(\infty)]$ even in the simple epidemic graph model above is $\#P$ -hard [69], corresponding to counting the number of solutions to an NP-complete problem. Even computing the probability that a particular node in a graph will get infected is $\#P$ -hard [70]. No known method of computing a $\#P$ -hard value in polynomial time is known, and moreover it is very unlikely that such a method exists [71].

As a consequence, there is no known computationally tractable way to determine the expected final size of an epidemic from R_0 as defined here, nor from any other definition of R_0 that is known to be polynomial-time computable.

Fortunately, for any public health policy purpose, the exact expected number of people infected need not be computed. A close approximation suffices. Indeed, we argue that it would be pointless to compute a value to within much greater precision than the accuracy of the input data. Theorem 3.2.1 assures that the sample average of multiple independent simulations well-approximates the expected number of infections with high probability. The proof, which relies on a concentration inequality, is given in Appendix C.

Theorem 3.2.1. *Let $G = (V, E)$ be the contact network of an epidemic model. For any $\epsilon > 0$ and $\delta > 0$, one can compute, in time bounded by a polynomial in n , a value \hat{R} such that*

$$P\left(\left|\hat{R} - E[F]\right| > \epsilon\right) < \delta.$$

For instance, suppose the actual expected fraction of the population to become infected is 0.15, according to the input data. Then a sample average of 3000 simulation trials is guaranteed to be precise within a factor of 1.09, *i.e.*, in the range $[0.137, 0.163]$ with probability more than 95%. There are similarly derived bounds on the required number of runs for other estimates such as the probability a particular individual or group is infected, or the distribution of outbreak size in a locality. (The method proposed in [66] can be proved

to produce a good approximation, along the lines of Theorem 1, but would require orders of magnitude more trials because it estimates the entire time stream, $R(t)$ for all t .)

One could hope that R_0 gives a coarse but useful approximation. One of the key features of the reproduction number is that it is supposed to have the following threshold property: when $R_0 > 1$ the epidemic will spread; when $R_0 < 1$ the epidemic will die out. To be precise, define R_0 to have the threshold property if the following is true: Introduce $n_0 = o(n)$ infected individuals to a disease-free population of size n ; if $R_0 > 1$, the expected number of people infected $E[R(\infty)] \geq \alpha n$ for some $\alpha > 0$, as $n \rightarrow \infty$; if $R_0 < 1$ then $E[R(\infty)] < \alpha n$ for all $\alpha > 0$ as $n \rightarrow \infty$. Equivalently, define F to be the expected fraction F of population that becomes infected, $F = E[R(\infty)/n]$. Then R_0 has the threshold property if

1. $R_0 > 1 \Rightarrow F = \Omega(1)$

and

2. $R_0 < 1 \Rightarrow F = o(1)$.

A counterexample to part 1 of the threshold property, that $R_0 > 1$ leads to a large epidemic, is straightforward to construct. For instance, one can just look at a line or grid graph as was done in [52]. In both of those, as we add nodes to the graph, the expected number of infected people is bounded above by a constant. Therefore F will go to 0.

We now assert a stronger property. Given a graph G and an epidemic on G with $R_0 > 1$, there is a graph \hat{G} such that the same epidemic on \hat{G} has larger reproductive number $\hat{R}_0 > R_0$ but smaller (relative) final epidemic size, i.e. $\hat{F} < F$. The proof is given in the Appendix.

Theorem 3.2.2. *Let $G = (V, E)$ be a graph. Let H be another graph obtained by connecting two copies of G with an extra edge (chosen appropriately). Suppose epidemic process on G with transmission probability $p \in (0, 1)$ has basic reproduction number R_0 and expected fraction of infected $E[F]$. Then the epidemic process on H with transmission*

probability p has basic reproduction number $R_0^H > R_0$ and expected fraction of infected $E[F^H] \leq \frac{1+p}{2}E[F] < E[F]$.

That $R_0 > 1$ is not a sufficient condition for a disease outbreak is a consequence of a more fundamental failure of a monotonicity property. Consider two networks with the same pathogen but different basic reproduction numbers R_0^1 and R_0^2 respectively (due to different mixing patterns). If $R_0^1 > R_0^2$, then one would desire that on average, the first epidemic would have a larger proportion of infected than the second one. However, this is not the case as illustrated by the following example.

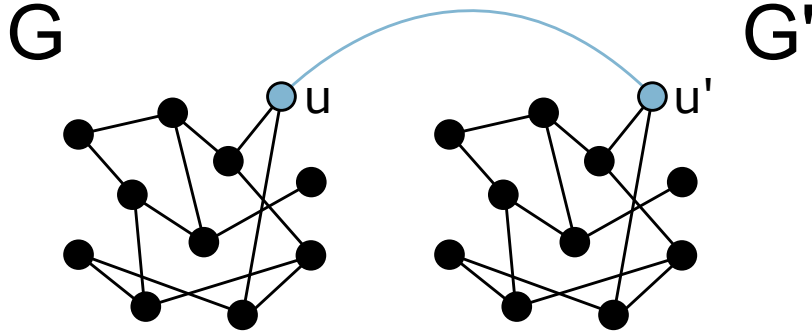


Figure 3.1: **Construction of a graph that has higher R_0 than any given graph G , but much lower F .** R_0 has increased, as for unchanged duplicated vertices it stays the same and we added an edge. F has decreased as the spread to G' is conditioned on traversing the edge (u, u') .

Suppose we have a graph G with basic reproduction number R_0 . Construct a new graph G' by taking two copies of G and connecting them via an edge at some node. See Figure 3.1.

Now we have two copies of the original graphs plus an extra connection. This makes R_0 of the new graph slightly higher than that of the original one. However, the expected fraction of people has decreased as invading both copies relies on transmission through the new edge. Details are given in the Appendix.

Next, we prove that no finite value of R_0 ensures part 1 of the threshold property. For any R , however large, and any $\epsilon > 0$, however small, there exists a graph for which $R_0 \geq R$ and $F \leq \epsilon$.

Theorem 3.2.3. *Let $G = (V, E)$ be a graph. Let H comprise m copies of G connected by $m - 1$ extra edges (chosen appropriately). Suppose epidemic process on G with transmission probability $p \in (0, 1)$ has basic reproduction number R_0 and expected fraction of infected $E[F]$. Then H has edge density slightly more than G , yet the same epidemic process on H has basic reproduction number $R_0^H > R_0$ and expected fraction of infected $E[F^H] \leq \frac{1+p}{m(1-p)}E[F]$. Moreover, $R_0^H \rightarrow \infty$ as $m \rightarrow \infty$.*

The choices of the extra edges and the necessary calculations are given in the Appendix. Lest the reader be concerned that the construction in Theorem 3.2.3 may be pathologically unrealistic, Figure 3.2 displays a real population distribution whose topology is similar to that of H .

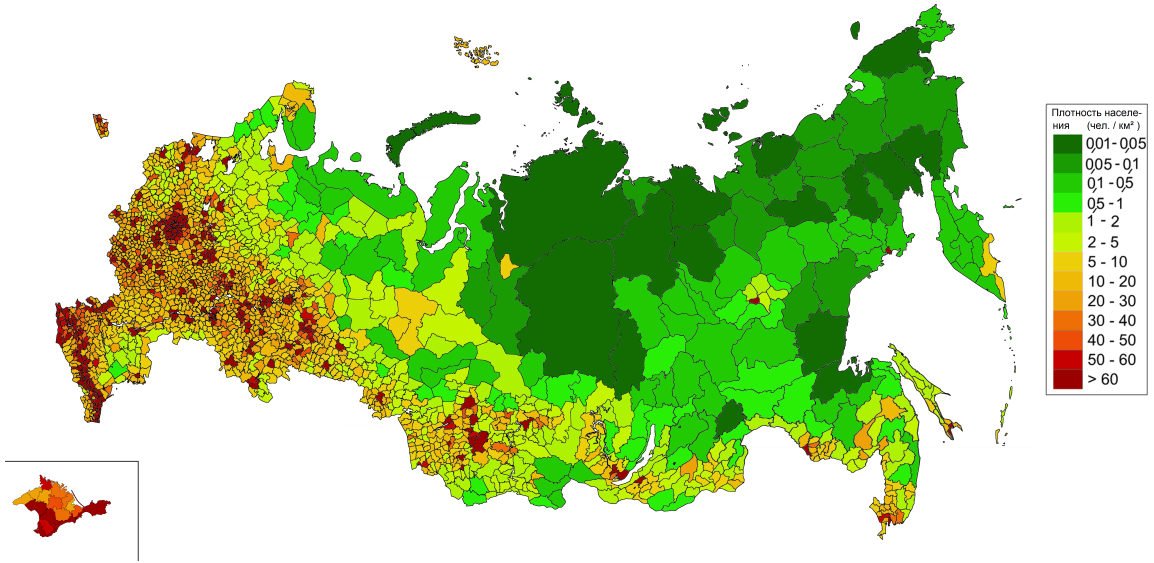


Figure 3.2: **Example of a real population distribution along a line.** The Russian population in its Asian regions closely follows the Trans-Siberian Railway. The numbers in the legend are densities in units of people per square kilometer ([72]).

The counterexample to part 2 of the threshold property, that $R_0 < 1$ ought to imply a negligible epidemic, is a bit more involved.

Theorem 3.2.4. *Let $R_0 < 1$. There exists a sequence of graphs $G_n = (V_n, E_n)$ with $|V_n| \geq n$, $p^* \in (0, 1)$ and $F > 0$, such that epidemic processes with transmission probability p^**

have basic reproduction numbers $R_0^n \leq R_0$, and their fractions of population infected, F_n , satisfy $E[F_n] \geq F$ for all n .

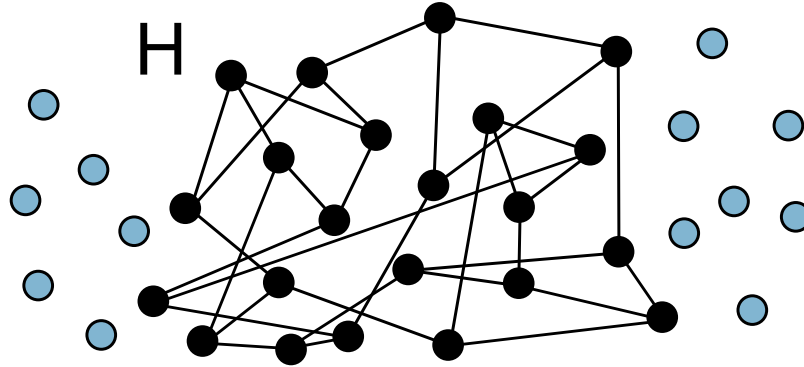


Figure 3.3: **Construction of a graph that for arbitrarily low R_0 has constant F .** This is a member ($n = 28$) of a sequence of graphs with arbitrarily low fixed R_0 that always have at least F fraction of infected. H is a well-connected central component with constant degree vertices (to keep R_0 from growing). A bounded fraction of blue vertices is added to H to get R_0 as low as required.

The construction relies on creating a sequence of graphs with a large, relatively densely connected component and weakly connected/disconnected outlying nodes. The central component has to be dense enough to make the spread relatively likely but not so dense that R_0 grows arbitrarily large. In our construction, all vertices in the connected component have the same degree and relatively uniform structure, giving the so-called expander property. See Figure 3.3.

Calculations in the Appendix prove that for any fixed $\rho > 0$, our constructed contagion network has basic reproduction number $R_0 = \rho$, but has an expected outbreak size proportional to the population size. In other words, as the population size $n \rightarrow \infty$ the fraction of infected $F_n \rightarrow F > 0$.

Theorems 3.2.3 and 3.2.4 prove that the basic reproduction number can wildly overestimate or underestimate the final size of the epidemic. It can be wrong by more than any constant factor in predicting the size of the epidemic and, furthermore, cannot even forecast whether there will or will not be an epidemic in the first place. This concludes our theoretical critique of applying R_0 to contact networks.

3.3 Conclusions

Our theoretical analysis demonstrates that there are epidemic processes for which $R_0 \geq 1$ is neither a sufficient nor necessary condition for a large outbreak. Furthermore, R_0 should not be used to estimate the size of the outbreak, for it can perform arbitrarily poorly as an approximation. According to Theorem 1, it can vastly overestimate the extent of an outbreak: there are cases with arbitrarily large R_0 , but in which asymptotically 0% of the population is exposed. According to Theorem 2, it can vastly underestimate the danger of an outbreak: for any $\epsilon > 0$, there are cases with $R_0 < \epsilon$ but in which the expected number exposed grows in proportion to the population size. These results explain and extend criticisms of R_0 reported in the literature.

As inhomogeneous network models replace homogeneous models of disease transmission, there cannot be a perfect replacement for the reproduction number. This is because exact computation of the expected fraction infected is inherently too complex ($\#P$ – hard). We have proved that the sample average of fraction infected is a close approximation to the expected fraction infected, with high probability. Therefore we propose that the sample mean should be used instead of R_0 . This computationally practical measure will provide more accurate forecasts of disease spread. Hence it will help identify the most threatening disease outbreaks, and evaluate the impacts of different potential interventions.

CHAPTER 4

CASE STUDY OF THE YEMEN CHOLERA EPIDEMIC

4.1 Introduction

Poor worst-case performance of a mathematical procedure does not necessarily imply poor behavior in practice. It could be that our theoretical analysis depends on rare or unrealistic cases. This calls for an empirical analysis to complement the theoretical one. Computational studies in the literature that investigate deficiencies of R_0 have been on small to medium scale real networks and various artificial networks. We perform a large scale case study on the present cholera epidemic in Yemen. The contact network model has more than 25 million nodes and 10^{13} edges. We find that R_0 is quite inaccurate as a predictor of the eventual disease spread over a range of data estimates. Furthermore, focusing on R_0 can mask important factors that affect the spread of the epidemic.

4.2 Case study

The examples in the previous Chapter were carefully constructed to get the desired properties. Perhaps real-world contact topologies do not exhibit such aberrant behaviour. To investigate whether the theoretical failings of R_0 extend to practical situations, we have developed a model of the current cholera outbreak in Yemen. Our model is substantially larger than other networks models for which R_0 has been scrutinized.

Destruction of infrastructure and sanitation systems led to the start of a cholera epidemic in October 2016 [73]. It seems that before appearing in Yemen, this originally South Asian strain caused outbreaks in East Africa [74]. Since then, it has become the largest cholera epidemic in recent history. There have been over a million reported cases and more than 2000 deaths.

The outbreak involves a large population that includes heterogeneities relevant to disease transmission: population density, socioeconomic differences, and geographic features. Therefore, this is a case where R_0 might perform poorly as a metric for epidemic description, as opposed to a small uniform population at a school or workplace.

We investigate whether we can use the basic reproduction number to inform us about the size of the epidemic, perhaps not exactly but at least with some accuracy. Another property we investigate is whether relying on the reproduction number as a major summary statistic obscures other disease parameters that play a more important role in describing the epidemic.

4.2.1 Model overview

We built the contact network model based on detailed demographic data. Each node in the network represents either a person or a water source. Nodes representing people are strongly connected to their local water sources. They are also weakly connected to distant water sources representing the possibility of long-distance travel. Water sources are connected between each other by two-way edges if they represent reservoirs or wells, and by one-way edges if they represent naturally flowing water, like creeks and rivers.

The disease progression in people follows the SEIR model. Susceptible individuals can get infected through exposure to *Vibrio cholerae* either by drinking contaminated water or eating contaminated food. They then become exposed, infected but not infectious while the disease incubates. Afterwards they progress to an infectious stage, during which they can add infectious bacteria to the local water source. Finally, they become removed from the population under consideration by either recovering or dying. See Figure 4.2.

Water sources may be contaminated in two ways. People who are infectious can contaminate water sources. Also, bacteria can seep to a nearby reservoir or well, or to a downstream river. The flow of the epidemic is simulated as a stochastic process on the network with each time step taking a day.

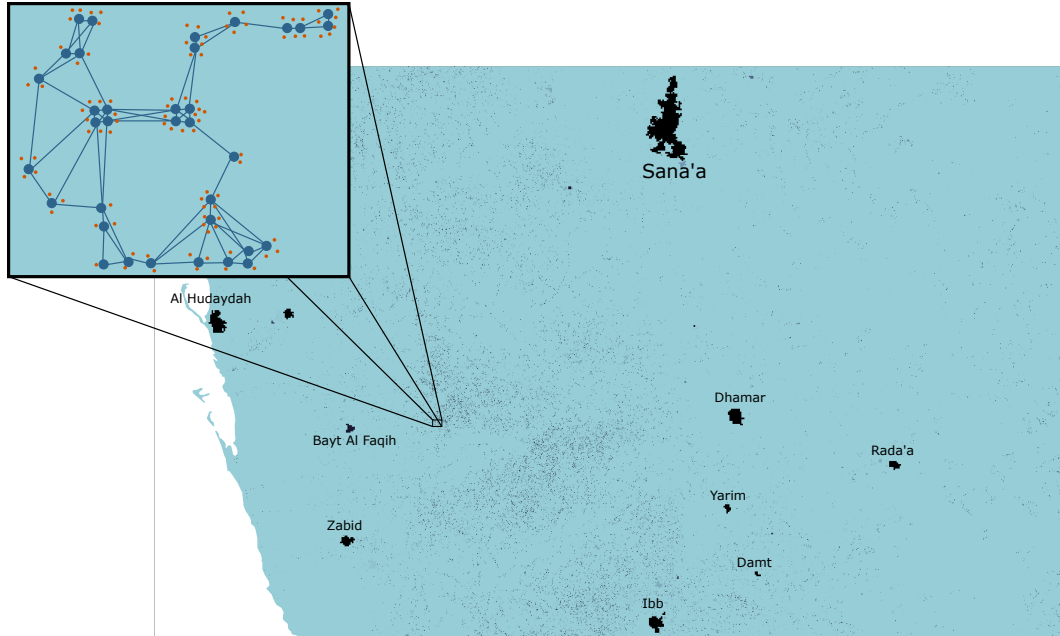


Figure 4.1: **Contact network** Example of the contact network for a an approximately 4 km² region of Raymah Governorate. Blue nodes and links represent water sources and their connections. Orange nodes represent people.

4.2.2 Results

The model was calibrated against WHO surveillance data at the country level [75]. We used the data up to February 2nd 2018, the final date of detailed WHO reports. Overall, a large fraction of simulation sample paths match the historical pattern of the epidemic. See Figure 4.3. Details on data sources, parameter values, network construction, simulation and validation are given in Section 4.3.

The basic reproduction number, calculated from simulations, was 3.05. This is within general bounds estimated for previous outbreaks. However, as with other outbreaks, R_0 does a poor job at summarizing the scale of the epidemic in Yemen. Knowing R_0 does not seem to add any extra information. One also would be hard pressed to make any predictions about the outbreak sizes based on the the relative magnitude of R_0 . See Table 4.1.

In our simulations, R_0 is sensitive to some parameters in a representative manner. For instance, significantly increasing the rate of water source access in the model results in a

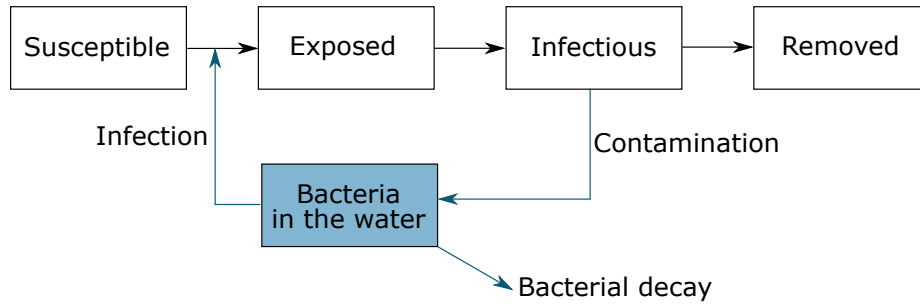


Figure 4.2: **Diagram of the cholera model.** People can be Susceptible to cholera, Exposed, Infectious or Removed (immune to infection and not part of the epidemic process). Black arrows represent state transitions. Susceptible people get infected by exposure to bacteria in the water. After an incubation period, they transition from Exposed to Infectious state. Cholera bacteria are shed into the water supply by Infectious people. There is a natural rate of bacterial decay that removes them from the environment.

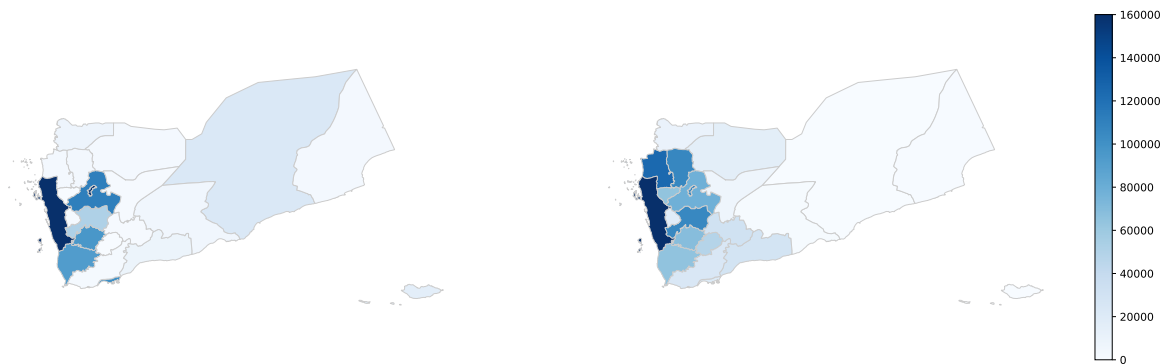


Figure 4.3: **Example of a simulated (left) vs reported (right) disease spread.** Large numbers of infections occur in densely populated urban areas around the capital of Sana’a and western parts of the country. Due to the model not accounting for population movements, the simulated epidemic has fewer infections in the main refugee destinations: Hajjah and Amran.

significant increase in R_0 and correspondingly significant increase in the number of infections. However, R_0 is almost completely insensitive to the effect of other parameters. See Figure 4.4. All parameters of the model were kept the same except for the probability of contaminating a distant water source (via long distance travel). Doubling this parameter caused a negligible change in R_0 but resulted in more than a factor of 5 increase in F . This aligns with genomic evidence suggesting that recent outbreaks start from a single introduction followed by continuous local spread [79, 74]. This indicates that R_0 does not have the sensitivity one would require from a good summary statistic.

Table 4.1: **Cholera outbreaks, their R_0 and affected population**

Outbreak	Cases (thousands)	R_0	F	Source
1. Angola 2006	82	5.90	0.41%	[76]
2. Bangladesh 1991	235	4.40	0.22%	[77]
3. Yemen 2016	1055	3.05	3.73%	[75]
4. Haiti 2010	809	1.55	8.1%	[78]
5. Zimbabwe 2008	99	1.15	0.80%	[57]

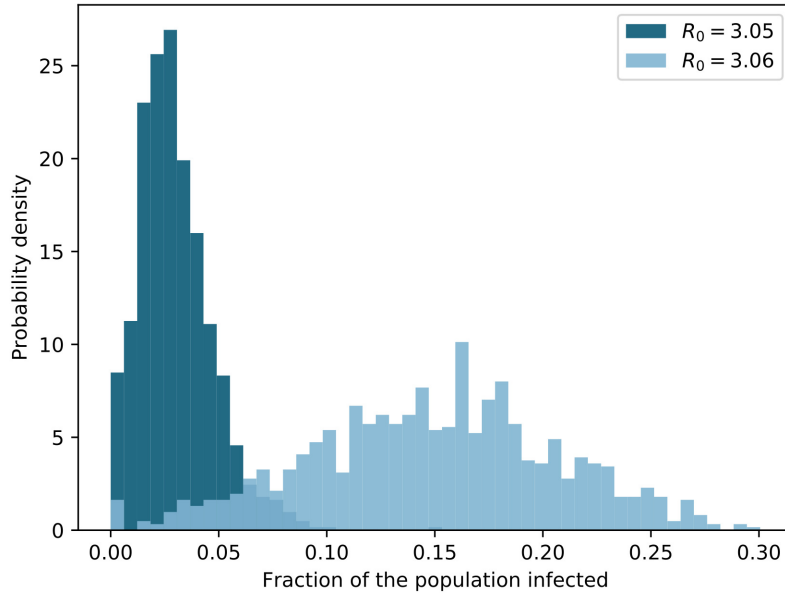


Figure 4.4: **Final epidemic sizes for varying probabilities of distant infections.** Increasing probability of distant infection by a factor of 2 results in a significant increase in the final epidemic size but has negligible impact on R_0 .

4.3 Cholera Outbreak Model Specifications

Section 4.2 has given an overview of our simulation model and its validation tests. This section specifies our model, data sources, and validation tests. We intend to provide sufficient detail here to enable other researchers to replicate our results.

4.3.1 Contact network

We created a large directed graph to represent Yemen. The network has two types of nodes: people and water sources. There are 24 million people nodes. A water source node represents a well, river, or reservoir that could be used for drinking water.

The water sources and their connections were created using data from [80] and [81]. We partitioned the country into a grid of 100m by 100m squares. Of these 5×10^7 squares, we retained those that have a river (based on [81]) or at least one person (based on [80]). We reasoned that any grid square that supports a nonzero population must have a water source. The network has a water source node for each retained grid square. There are 1,241,661 water source nodes of which 740,091 correspond to squares containing a river.

There is a directed arc from Water source u to water source v if all of the following conditions are met:

- u and v are within distance m_W , the maximum water spread radius.
- There is no water source node w between u and v .
- If both u and v are river nodes, v has to be downstream of u .

People are placed into the grid squares based on data from [80]. People nodes have a strong two-way connection to the water source in their grid square. This represents people being able to contaminate their local water sources and get infected by accessing water from their closest water sources. People nodes also have weak one-way connections to all other water sources. This represents people traveling and contaminating distant water sources.

4.3.2 Disease progression and transmission

Each person node can be in of the four stages according to the SEIR model. The durations of disease stages are uniformly distributed random variables with ranges shown in Table

4.2. Number (concentration) of bacteria represents the level of contamination. Bacteria have a uniformly random lifespan B_d between 7 and 14 days.

Disease can spread each day in one of the following ways:

- Infected people add bacteria to their closest water source with probability p_h
- Infected people add bacteria to a distant water source with probability p_d . The distant water source is chosen among all water sources with probability proportional to the number of people at that water source.
- Bacteria in the contaminated water source infect healthy people adjacent to it with probability $p_w n_b / (K + n_b)$, where p_w is rate of contact with the water source, n_b is the concentration of bacteria in the water reservoir, K is the half saturation constant. K represents the concentration of bacteria that yields 50% of the maximum chance of infection.
- Bacteria at water source u can multiply and seep over to the nearby connected water source v with probability $p_s / d(u, v)$, where p_s is probability of spread and $d(u, v)$ is the distance between water sources.

4.3.3 Simulation

We initially introduce a random infected person within 10 km of Sa'ana'a city center [73]. This person starts in Infectious stage with randomly generated duration. The simulation runs are restarted until an epidemic occurs, *i.e.*, disease spreads to enough (> 100) people.

Conditioned on the epidemic occurring, the simulation continues to run one day at a time. Each day the disease spreads according to the stochastic process outlined above. At the end of each day people advance in their disease stage or enter new stages.

The simulation runs for 497 days, to mimic the time from the first reported cholera case in Sa'ana until the last day of detailed WHO reports. We terminate the simulation early if

Table 4.2: **Simulation parameters**

Parameter	Description	Value	Source
m_w	maximum radius of water contamination	1 km	calibrated
p_{hw}	maximum probability person gets infected by drinking contaminated water	8.0×10^{-2} day ⁻¹	calibrated
p_w	probability an infectious person infects local water source after contact	6.5×10^{-2} day ⁻¹	calibrated
p_d	probability an infectious person infects distant water source after contact	8.0×10^{-4} day ⁻¹	calibrated
K	half saturation constant	20 units	calibrated
p_s	multiplier for probability of bacteria spreading to connected water source	1.36×10^{-4} day ⁻¹	calibrated
E_d	duration of the exposed stage	1 – 5 days	[82]
I_d	duration of the infectious stage	4 – 7 days	[83]
B_d	lifespan of the bacteria	7 – 14 days	[83]

the disease dies out, *i.e.*, if there are no people in an Exposed or Infectious stage and there are no bacteria in any of the water sources.

4.3.4 Model limitations and validation

All available Yemen population data estimates appear to be extrapolated from 2004 Census. We ran the model both with the raw data from [80] and with data adjusted by newer estimates [84]. The adjustment was performed by proportionally scaling the number of people in each square for each governorate. We use the governorate boundaries from [85]. The calibrated parameters and final R_0 were only slightly sensitive to the adjustment (population was 16% larger, final R_0 increased by 9% due to the higher density created by scaling) and the overall qualitative results were the same.

The most comprehensive available waterway dataset is in [81]. It appears to be complete, but does not distinguish among intermittent, ephemeral, and permanent streams. However, the alternative waterway data sources are incomplete and have no evidence of greater reliability. Therefore, we used the data from [81], and modeled all streams as perennial. Even moderately accurate data for other water sources is not available. Urban reservoirs and pipelines are in greatly varying states of functionality due to the war and

consequent governmental disarray. In rural areas, a large number of illegal wells have been dug because of significant drops in the water table, making accurate data inherently difficult to obtain. As a proxy for absent data, the model includes a water source node in every grid square with nonzero population, even if no well, river, pipeline, or reservoir is known to be present. The m_W parameter then approximately models the spread of bacteria to nearby squares.

The early WHO estimates for cholera cases could also have some issues. The monitoring stations were set up months after the start of the outbreak. Even afterwards, the country being an active war zone and general failure of the infrastructure makes monitoring extremely difficult.

Some potentially important factors that we do not model are interventions, population movements, and climate. There have been various small scope intervention efforts throughout the time period we consider. Detailed data for them is rather difficult to come by. There are some limited data on population movements based on internally displaced people reports [86], but Yemen's state of civil war makes more detailed and reliable reports highly unlikely to appear. Finally, cholera seems to have a seasonality element that is quite difficult to quantify [87]. Of these excluded factors, population movement turns out to appear significant. However, modeling the future movement of refugee populations is well outside the scope of an epidemiology study.

We validated the model by comparing simulation results with data that was not directly used to calibrate the model - the infections in different governorates. See Figure 4.5. Simulated epidemics follow patterns close to the observed one in most governorates that were relatively untouched by the civil war. As expected, large numbers of infections occur in densely populated areas connected to the capital (the origin of the epidemic). For instance, in the governorates of Al Hudaydah, Ibb, Taizz, and capital governorate Sana'a along with the city itself (Amanat Al Asimah), More than 50% of the sample paths land within 50% relative error of the actual outbreak. Nearly 10% of the sample paths are within 10%

relative error compared to the actual outbreak. Most sparsely populated or hard-to-reach governorates also exhibit realistic behavior. For instance, about 10% of the time the island of Socotra is not touched by the epidemic. However, if the disease manages to take hold in the island, one observes a small outbreak.

For some governorates the fit is quite poor. However, the governorates with poor fit match two major events of the conflict and refugee movements [88, 86]. The primary instance occurs in the governorates of Hajjah and Amran. These have been the most popular destination of refugees and they are the ones where the model most significantly underestimates infections. The other significant instance is the city of Aden, which was the main military point of contention during 25 March – 22 July 2015 and has been held by the Saudi-led coalition since then. The neighboring areas experienced a flood of refugees at the beginning of the conflict, making the model underestimate those numbers. Afterwards, there are at least two possible explanations for the overestimates in Aden: an overestimate of the population and an underestimate of the infrastructure level. The model assumes that infrastructure there is in the same state as in other cities. However, other cities are controlled by sides of the conflict with less funding than Aden's. These features of the conflict qualitatively explain the model's inaccuracy.

There is large amount of variability in the model. As with all large scale epidemic models, it is impossible to know whether this is inherent to the actual process or just a feature of the model. Only one sample path of the epidemic can be observed in the real world, making any estimate of the variance statistically invalid.

4.4 Conclusions

Our case study of the Yemen cholera epidemic demonstrates that R_0 has very limited practical applicability. The reproduction number does not help predict the size of the outbreak. It tells us virtually nothing as a general summary statistic. Finally, it is not sensitive to some critically important epidemic process parameters. The failures of R_0 we observe are confir-

matory of deficiencies previously reported in empirical studies [65, 66, 28, 67]. Compared with prior studies of real disease outbreaks, the one reported here is, built on a much finer geographic scale (cell size $100m \times 100m$), so as to capture the propagation and movement of bacteria in water sources. The resulting population scale (average cell population 19) is considerably finer as well.

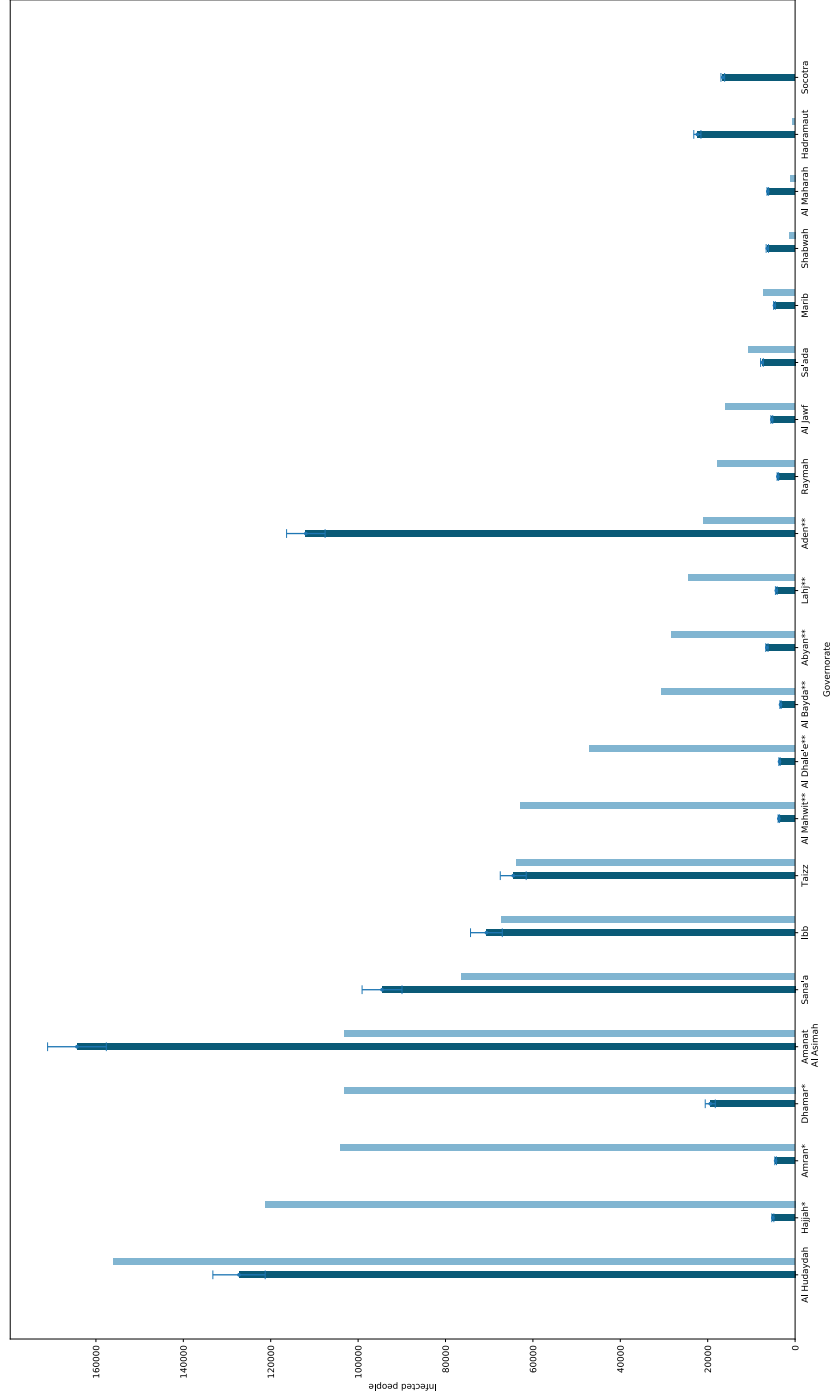


Figure 4.5: **Plot of estimated infections by governorate.** The numbers reported by WHO through February 2nd 2018 [75] are compared to model projections (median with 95% CI) for the same period. *Northwestern governorates that were the main refugee destination. **City of Aden and surrounding governorates that experienced heaviest fighting and initial refugee flood.

CHAPTER 5

EPIDEMIC CONTROL

5.1 Introduction

Throughout history, infectious diseases have been some of the most damaging phenomena faced by human societies. In previous chapters we discussed how to measure the effect of a potential outbreak. Having this information helps public health organizations to decide whether there is a need for preventive intervention [26]. However, this still leaves the question of how best to carry out such an intervention. Sometimes, this boils down to ensuring access to the vaccines. Other times, e.g. for quarantine-based solutions, it becomes a question of prioritizing some populations over others. Good disease prevention strategies are an important problem in public policy. Furthermore, they inspire interesting mathematical optimization challenges which we study in this chapter. We explore the problem of minimizing the effect of a spreading process on a network subject to an intervention budget. Although we are mostly interested in stopping disease spread, this framework can be applied to other spreading processes, including malware on an IT network, misinformation on social networks, and illegal items on transportation networks.

There has been quite a lot of work done on minimizing the effect of spreading processes on networks, by researchers with backgrounds ranging from epidemiology to physics [89, 90, 91]. We will summarize the main approaches here. For more detailed reviews we direct the interested reader to a few recent surveys [92, 29]. In the context of a network, interventions can be classified into three groups: node removal, edge removal, or decrease in the probability of transmission. Depending on the network resolution, node removal can correspond to vaccination [93], treatment [94], case isolation or sometimes quarantine [95]. Quarantine, along with social isolation (e.g. school closure) can also be viewed as an

intervention based on edge removal [96]. For instance, if network nodes correspond to individuals, quarantine can be represented as an edge cut between subsets of a larger population [97]. Finally, there are intervention strategies that reduce the probability of infection, without completely cutting off the contacts. These mostly take the form of behavioral changes aimed at either diminishing interaction frequencies or reducing the transmission risk of each interaction. Some examples of the latter are promotion of hygiene such as hand washing, boiling water, or cooking food to the right temperature, and barrier precautions such as masks, condoms, latex gloves, etc [98]. These intervention strategies can be distributed over the network uniformly and randomly, or aimed at specific parts of the network. In this chapter, we will focus on the targeted interventions. We won't consider measures that reduce probability of infection as they are uniform in their nature. If we allow different nodes and edges to have different costs and objective value coefficients, then one can easily transform from one type of problem into another. We picked node removal as it more naturally corresponds to the most common prevention strategy - vaccination.

The ultimate goal of any intervention is to minimize the number of victims of a disease. However, even evaluating this objective is not a simple task. As we discussed in Chapter 3, we can approximate the expected number of infected but cannot evaluate this number with a simple expression. Therefore, many papers in the literature [29] instead focus on minimizing a proxy that hopefully would lead to minimizing the impact of the disease. One common approach is to minimize a metric seen as a threshold for the disease, often R_0 . However, this task itself is also computationally difficult. Recall the definition of R_0 for a spreading process discussed in Section 3.2:

$$R_0 = \frac{1}{|V|} \sum_v p\delta_v$$

As we will see in Section 5.3, minimizing this metric via node removal is NP-Hard. Moreover, it is impossible to approximate within any constant multiplicative factor unless P=NP.

Hence, most approaches resort to greedy heuristics, based on removing nodes with high degree [30, 99, 100]. For very large graphs, some authors resort to doing this without reevaluation. This approach can be quite useful in the absence of global information on a graph [31] as it provides a reasonable heuristic for spending limited removal budget on nodes that are hopefully of high importance.

Another class of intervention strategies is based on various centrality metrics. Although node degree can be a good indicator for the relative importance of a node in terms of its spreading capabilities, it can be inaccurate [30]. To alleviate that, one should consider broader impact of a node. For instance, a vertex that has many high-degree neighbors is more important than its degree alone would indicate. This gives rise to various node centrality measures. Perhaps the most common ones are betweenness centrality and eigenvector centrality, although there are others [101]. For a given graph $G = (V, E)$, with an adjacency matrix A , we can define betweenness centrality $C_B(v)$ of a node v as:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}},$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v . The eigenvector centrality C_E of the vertices is defined as a solution of the equation:

$$AC_E = \lambda C_E,$$

Where λ is the largest eigenvalue of A . Different authors have used these centrality measures to inform a variety of interventions strategies [102, 103, 104]. Such interventions are also commonly used to minimize another proxy for the number infected: the aforementioned λ , the largest eigenvalue of the graph adjacency matrix. It has been shown that these approaches can perform arbitrarily badly for this purpose [105]. Nevertheless, they often empirically outperform interventions based on the vertex degrees [30].

Finally, there are many approaches based on dismantling the network [106, 107, 108]. The idea is to partition the network, hence confining an epidemic to a small component. Variants of the problem differ in objectives, and constraints on the size and number of components. Researchers have attempted both exact [109, 107] and heuristic approaches [106, 108] to these problems. We will not consider at these approaches in detail, as they are essentially ‘dual’ to the problem we have at hand. *I.e.*, partitioning approaches typically aim to minimize the number of removed nodes to achieve a good partition, rather than to achieve the best possible partition subject to a budget constraint. The former is more appropriate for network partitioning problems, as a small number of nodes cannot guarantee any kind of partition. In our context, one can view it as a limit of a general spreading problem when the probability of edge transmission approaches one. If the initial infected node is uniformly random, this becomes a problem of minimizing the sum of squares of network component sizes [107].

There are a few more broad areas of related research that we want to mention. Intervention efficacy can be affected by human behavior [110]. For instance, some individual could refuse to participate in intervention measures, there could be general change of behavior with the introduction of quarantine, etc. These issues are usually investigated in a game theoretic framework which is outside the scope of this work. There is also a significant amount of related research in the contexts of the firefighter problem [111] and network flow interdiction [112, 113]. However, both problems are usually considered in a deterministic setting and on graphs with very special structure, e.g. trees and grids.

In this work we want to establish truly optimal (rather than heuristic) interventions in a broad setting. The remainder of this chapter is organized as follows. Section 5.2 rigorously defines the problem of minimizing the effect of stochastic spread on a network under general conditions. It provides an exact albeit impractical IP formulation for this problem. It also describes how to obtain a more practical formulation using sample average approximation. In Section 5.3, we will discuss solution approaches. We study the

computational complexity of SAA for the problem and its approximability. We will also present a result on inapproximability of intervention based on R_0 . Finally, we show how SAA based approach is viable for small networks. We perform a series of experiments on real contact networks. In these experiments we evaluate our solution against degree and centrality-based interventions. We demonstrate that using our intervention can result in significantly better outcomes.

5.2 Problem statement

Consider a directed graph $G = (V, E)$. An edge $(v, u) \in E$ in the graph G represents the possibility that v may expose u to the disease. At any given time t , each node is one of the three states: Susceptible (S), Infectious (I), and Removed (R). Initially, during step $t = 0$, we draw one node to become infectious from some distribution P . All the other nodes are susceptible. For notational convenience, we will only deal with the case of one initially infectious node, though generalization to multiple initial infections is straightforward. Every infected node v stays infectious for some (possibly random) time. This duration of infectiousness can be a different distribution for each node. At the end of infectious period, the node becomes removed. Each infectious node v transmits infection to its neighbouring susceptible node u according to some stochastic process $\beta(u, v)$. If the transmission is successful, u becomes infected itself. The process stops when there are no more infectious nodes. Note that this is a significant generalization of the spreading process defined in Section 3.2. It can be difficult to parametrize such processes, though the recent availability of high quality contact data is making it increasingly possible [114, 115, 116].

We have a budget to immunize b nodes before the process starts (effectively rendering them removed). The goal is to minimize the total expected number of infections after immunization. We assume that immunization is perfectly effective, i.e. there is no possibility that a vaccine can fail. Furthermore, we assume that our intervention is independent of the spreading process. For instance, this would mean that a vaccination campaign would not

introduce significant behavioural changes in the population. Finally, we assume we have full information about the spreading process, i.e. we don't have to learn and estimate the network topology or the transmission process.

Dealing with the problem in this form is quite difficult. In particular, it is very difficult to establish the effects of an intervention. However, we can look at it from a different point of view. Observe that any realization of our spreading process is equivalent to a rooted subgraph H of G . The root of H corresponds to the initially infectious vertex. Every edge in H corresponds to a possible transmission event. I.e. an edge $(u, v) \in H$ corresponds to the sample paths where u could have transmitted the disease to v . We consider these paths regardless of whether the transmission actually occurred e.g. if v was infected by some other vertex first. Notice that these subgraphs, from now on called scenarios, partition the whole probability space. Furthermore, by independence of the intervention and the spreading process, we can observe the effect of the intervention by simply removing the required vertices from each scenario and performing a search from the root (if it wasn't immunized). This allows us to write the following (exponentially sized) integer formulation:

Sets:

- V - nodes
- E - directed arcs
- $S = |V| \times 2^{|E|}$ - possible scenarios
- E_s - set of all active edges in scenario s , $s \in S$

Parameters:

- b - immunization budget (in nodes)
- p_s - probability of scenario s , $s \in S$
- v_s - root node in scenario s , $s \in S$

Variables:

$$\bullet d_v = \begin{cases} 1, & \text{if node } v \text{ is immunized} \\ 0, & \text{otherwise} \end{cases}$$

$$\bullet x_{sv} = \begin{cases} 1, & \text{if node } v \text{ is infected in scenarios } s \\ 0, & \text{otherwise} \end{cases}$$

Formulation:

$$\begin{aligned} \min \quad & \sum_{s \in S} p_s \sum_{v \in V} x_{sv} \\ \text{s.t.} \quad & \sum_{v \in V} d_v \leq b \\ & d_{v_s} + x_{sv_s} = 1 && \forall s \in S \\ & x_{sv} + d_v \geq x_{su} && \forall (u, v) \in E_s, s \in S \\ & 0 \leq x_{sv} \leq 1 && \forall s \in S, v \in V \\ & 0 \leq d_v \leq 1 && \forall v \in V \\ & d_v \in \mathbb{Z} && \forall v \in V \end{aligned}$$

Although technically fully specified, there are two problematic issues with this formulation. The first is that the formulation has size exponential in the number of edges. The second issue is that for a general spreading process it is difficult to calculate the probability p_s of a given scenario s . However, we can alleviate both issues by employing the sample average approximation (SAA) technique [117]. Instead of looking at all scenarios, we define S as a sample of N scenarios. This allows us to get the expectation with respect to the empirical distribution. Therefore, we have the following SAA Intervention Problem (SAAIP):

$$\begin{aligned}
\min \quad & \frac{1}{N} \sum_{s \in S} \sum_{v \in V} x_{sv} \\
\text{s.t.} \quad & \sum_{v \in V} d_v \leq b \\
& d_{v_s} + x_{sv_s} = 1 && \forall s \in S \\
& x_{sv} + d_v \geq x_{su} && \forall (u, v) \in E_s, s \in S \\
& 0 \leq x_{sv} \leq 1 && \forall s \in S, v \in V \\
& 0 \leq d_v \leq 1 && \forall v \in V \\
& d_v \in \mathbb{Z} && \forall v \in V
\end{aligned}$$

By the (Strong) Law of Large Numbers, the optimal value and solution of SAAIP converge to the optimal value and solution of the original problem with probability 1 as $N \rightarrow \infty$. Furthermore it does so exponentially fast with the increase of the sample size [117].

5.3 Hardness and inapproximability

Now that we have a workable formulation ??, the question becomes, how do we solve it? In this section we derive a few results attempting to answer this question. On the theoretical front, we prove that the decision version of SAAIP is NP-hard. However, a lot of optimization problems are NP-hard, and yet are routinely solved in practice. One indicator of some NP-hard problems being easier than other is their approximability. Therefore, we carefully investigate how well can we approximate a solution for SAAIP. Finally we will briefly discuss a few practical considerations.

Proposition 5.3.1. Define decision version of SAAIP: Given a set of scenarios S and a budget b is there an intervention with expected number of infected less than M ? This

problem is NP-Hard.

Proof. We reduce from the Clique Problem. The proof is similar to [118].

Consider a clique problem: does a given (undirected) graph $G = (V, E)$ have a clique of size k ?

Construct an instance of SAAIP with a graph $(V'_s \cup V'_a \cup V'_n, E)$ as follows. Let the budget $b = |E| - \binom{k}{2}$ and the target number of infected be $1 + \binom{k}{2} + Ck$, where $C = |V|^2$. Create $N = (|E| - \binom{k}{2})(1 + |E| + C|V|)$ scenarios. These scenarios all have a different starting vertex $s \in V'_s$ but share all other vertices. For each edge $e \in E$ we create a vertex in V'_a and for each vertex v in V we create C vertices in V'_n . In each scenario, the starting scenario vertex s has an edge (s, e) for all $e \in V'_a$. For each edge $e = (u, v)$ in the clique problem graph, we connect a vertex $e \in V'_a$ to all vertices in V'_n corresponding to u and v . See Figure 5.1

(\implies) Suppose there is a clique of size k in G . Immunize $|E| - \binom{k}{2}$ vertices in V'_n corresponding to the edges not in the clique. Then in each scenario, we have the following infected vertices: the starting vertex, the vertices corresponding to the edges of the clique, and the vertex groups corresponding to the vertices in the clique. Overall we have $1 + \binom{k}{2} + Ck$ infected in each scenario and hence the required objective value.

(\impliedby) Suppose there is a way to immunize $|E| - \binom{k}{2}$ to get an objective value of at most $1 + \binom{k}{2} + Ck$. Consider an optimal immunization.

The number of infected nodes without immunization is $(1 + |E| + C|V|)$. If we immunize m scenario starting vertices we reduce the objective value by

$$\frac{m(1 + |E| + C|V|)}{(|E| - \binom{k}{2})(1 + |E| + C|V|)} = \frac{m}{(|E| - \binom{k}{2})} < 1,$$

and don't affect any of the other scenarios. Likewise, immunizing any of the vertices in V'_n reduces the objective value by 1 and doesn't affect the infection status of any other vertex.

Suppose we immunized fewer than $|E| - \binom{k}{2}$ vertices in V'_a . Then there would be at

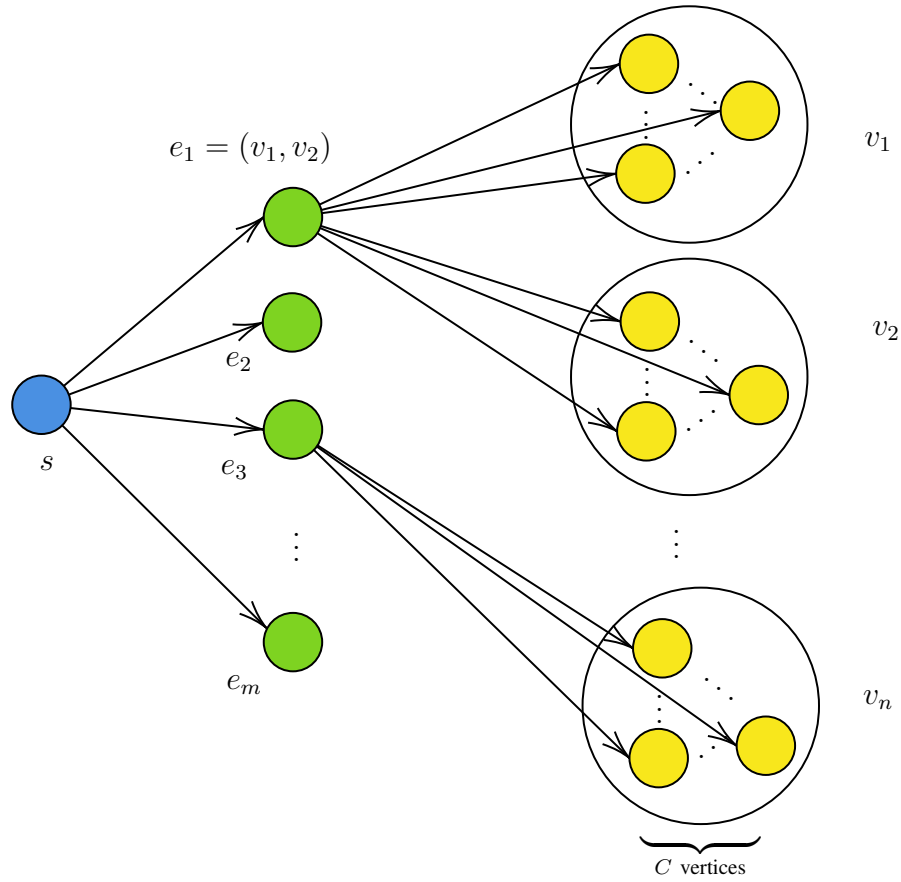


Figure 5.1: **Construction for the Proof of Proposition 5.3.1.** s is a unique starting vertex for each scenario in S . Vertices e_i correspond to the edges of the graph in the Clique Problem. Vertex groups v_j correspond to the vertices the Clique Problem

least $\binom{k}{2} + 1$ non-immunized vertices in V'_a . Each of them corresponds to an edge in E . In any graph, there have to be at least $k + 1$ vertices adjacent to $\binom{k}{2} + 1$ edges. Therefore, we haven't blocked the infection from reaching $k + 1$ clusters in V'_n . I.e. our objective value is at least

$$\left(1 + \binom{k}{2} + C(k + 1)\right) > 1 + \binom{k}{2} + Ck$$

which would contradict the optimality of our intervention. Therefore, all $|E| - \binom{k}{2}$ vertices are immunized in V'_a . The remaining $\binom{k}{2}$ non immunized vertices left are in V'_a and they are connected to at most Ck vertices in V'_n . I.e. we have at least $\binom{k}{2}$ edges in the original graph G connected to at most k vertices. The least number of vertices with $\binom{k}{2}$ is at least k . Therefore, we have a subgraph of G with k vertices and $\binom{k}{2}$ edges, i.e. a k -clique as required. \square

5.3.1 Approximability

Approximation algorithms often produce good results for NP-hard problems. In the next few theorems we show the difficulty of approximating SAAIP. Most approximation techniques are based on greedy or local search, randomized rounding, or a primal-dual schema [71]. We first demonstrate that the greedy heuristic can have an arbitrarily large gap. To deal with the latter two approaches we look at the gap between the integer formulation of the problem and different relaxations. Performance bounds for both randomized rounding and primal-dual schema depend on this gap, and large gaps would indicate that these approaches might not work.

Proposition 5.3.2. Intervention produced by a greedy algorithm for SAAIP with n nodes can be $\Omega(n)$ worse than the optimal intervention.

Proof. Construct an instance of SAAIP with a graph (V, E) as follows. Let the budget $b = 2$. Suppose we have n scenarios, each with a unique starting vertex. Each starting vertex is connected to vertices v_1, v_2, v_3 , and v_4 . v_1 and v_2 each have two children. v_3 and

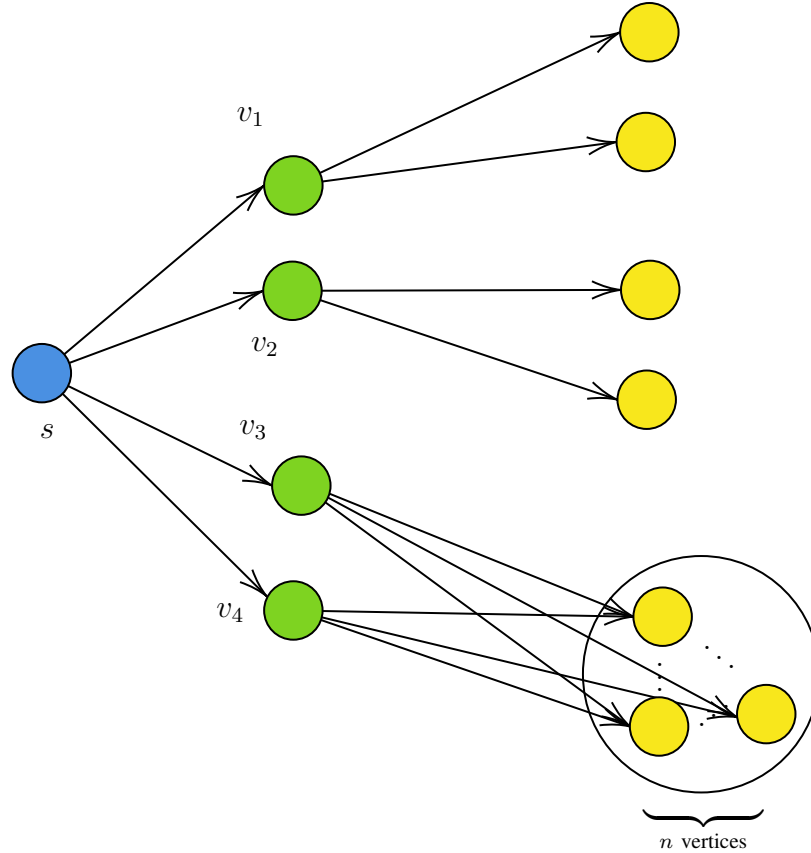


Figure 5.2: **Construction for the Proof of Proposition 5.3.2.** s is a unique starting vertex for each scenario in S . Vertices v_1 and v_2 have two successors each. Vertices v_3 and v_4 are connected to n vertices.

v_4 are connected to n vertices. See Figure 5.2.

Then the greedy heuristic would immunize vertices v_1 and v_2 , resulting in the expected number of infected $n + 3$. On the other hand, if we immunize vertices v_3 and v_4 , there will be 7 infected. Their objective ratio is at least $(n + 3)/7 = \Omega(n)$. \square

Next we look at the LP relaxation. Along with its application for the design of approximation algorithms, the gap between the IP and LP relaxation can be indicative of the IP solver performance. Regrettably, linear relaxation gives very little information about the actual solution of SAAIP.

Proposition 5.3.3. LP relaxation SAAIP ?? has an integrality gap of $\Omega(b)$.

Proof. Construct an instance of SAAIP with a graph $(V_s \cup V_1 \cup V_2, E)$ as follows. We

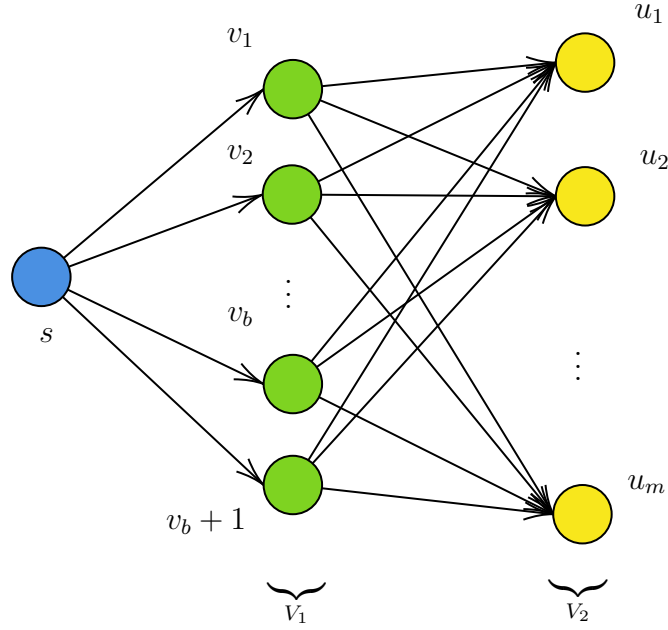


Figure 5.3: **Construction for the Proof of Proposition 5.3.3.** s is a unique starting vertex for each scenario in S . s is connected to vertices $\{v_1, \dots, v_{b+1}\} = V_1$. Each vertex in V_1 has an arc to each vertex in V_2 . By construction, with a budget b we cannot save more than b vertices.

have a budget b and n scenarios, each with a unique starting vertex. Each starting vertex is connected to $b + 1$ vertices in v_1 , which are in turn connected to m vertices in V_2 . Suppose $n \gg m + b$. See Figure 5.3.

Hence any immunization strategy can immunize at most b vertices, leaving $z = 2 + m$ infected. On the other hand, in the LP we can fractionally immunize each of the vertices in V_1 with $d_v = b/b + 1$. This leads to a solution with an objective value of

$$z_{LP} = 1 + \frac{1}{b+1}(b+1+m) = 2 + \frac{m}{b+1}.$$

For large m the ratio $z/z_{LP} \approx b$ which can be chosen to be arbitrarily large. □

One additional point we would like to make is that designing targeted intervention based on minimizing a threshold such as R_0 is also NP-Hard. In fact it is hard to approximate within any factor. The following proposition makes this statement precise.

Proposition 5.3.4. Intervention based on minimizing R_0 is NP-hard. Furthermore, it cannot be approximated to any constant factor unless $P = NP$.

Proof. Consider the spreading process as in Section 3.2. In this process, we have

$$R_0 = \frac{1}{|V|} \sum_v p \delta_v$$

p is just a constant. Therefore, (up to a multiplicative constant) the problem is equivalent to removing graph nodes to minimize the number of edges in the graph.

Suppose, for any graph $G = (V, E)$ and a budget b , we could solve the problem approximately within a factor of α . I.e. we could find an intervention with the basic reproduction number \hat{R}_0 , s.t.

$$R_0^{OPT} \leq \hat{R}_0 \leq \alpha R_0^{OPT}$$

However, for some budget b , $R_0^{OPT} = 0$ and hence \hat{R}_0 . Doing binary search over values of b , we would obtain minimal b s.t. removing b vertices removes all the edges. I.e. we found the minimal number of vertices that are adjacent to all the edge or the vertex cover. However, the latter is NP-Complete, and hence we cannot approximate optimal R_0 intervention within any constant factor unless $P = NP$. \square

5.3.2 Discussion and practical performance

We would like to make an observation about the approximability of the intervention based on R_0 . Although it is hard to approximate the optimal R_0 intervention within a constant factor, it is not hard to solve the same problem with differently specified objective. Minimizing the number of edges remaining after vertex removals is equivalent to maximizing the number of edges removed. However, the latter problem does admit a variety of approximation algorithms. The simple greedy heuristic gets within a factor of $1 - 1/e$ [71]. Rounding schemes can get an even better approximation ratio of $3/4$ [119]. So, in a way the problem is not as hard as it looks. Empirically, modern solvers can tackle it quite well.

The Gurobi 8.1 [120] software found optimal (within 0.1%) solutions on graphs with less than a 1000 nodes and any budgets without trouble.

For optimal immunization the picture is different. Our analytical results have been negative. Nevertheless, in practice modern solvers can tackle the problem for small to moderate numbers of scenarios with reasonable graph sizes. We will present detailed experimental results in the next section. We have been able to solve instances with up to 1000 scenarios on 200 nodes. Furthermore, greedy and local search based heuristics also perform reasonably well on these instances. However, the most popular decomposition approaches, e.g. L-shaped method [121] did not perform too well. Straightforward implementation doesn't work as well as one would hope. Although we can solve sub problems extremely efficiently (they boil down to calculating a network flow), the master problem does not converge quickly. In fact, in our experience, solving the full problem on a modern solver significantly outperformed such approaches.

5.4 Experiments

In this section we compare intervention based on solving SAAIP to a few popular alternatives. The alternatives we consider are: the intervention that minimizes R_0 , the intervention based on removal of nodes with highest betweenness centrality, and the intervention based on removal of nodes with highest eigenvector centrality. For SAAIP and the problem of minimizing R_0 , we solve the problem using Gurobi 8.1 [120]. For centrality based interventions, we employ the greedy approach used in the literature [30] of identifying and immunizing b nodes with highest centrality. Note that all the heuristics we described so far can be optimal in special cases. For instance, as we have already mentioned graph shattering can be optimal as the probability of transmission approaches 1. At the other extreme, degree-based interventions are also optimal when the probability of transmission is so small that we can ignore second order interactions. Between those, the centrality-based measures can work extremely well when we need to consider only relatively short infection paths.

The experiments use three real contact networks: Zachary’s karate club [122], Kenyan households contact network [123] and a French high school contact network [124]. Zachary’s karate club network is one of the most popular social networks in the network theory literature. The network consists of 34 members of a karate club, with links between pairs of members who interacted outside the club. It is considered to be a prototypical example of community structures in networks [125]. The second network consists of 47 people spread across 5 households in rural Kenya. These households were observed between April 24 and May 12, 2012, though the data correspond to only 5 days of continuous interactions. Each household member was equipped with a proximity sensor to record close (≤ 1.5 meter) interactions between them. A contact between two individuals was recorded if there was a packet sent in the last 20 seconds. The contact was considered continuous until no packets were exchanged for at least 20 seconds. We hope, this network is reasonably representative of interactions in a small rural community. Finally, the last network corresponds to contacts between 126 high school students in Marseilles, France. The dataset contains student contacts in three classes over 4 days in Dec. 2011. These were recorded using methodology similar to the one described above for the Kenyan village. This network should be much more representative of a well mixed urban population.

We used a general SIR model for disease progression. In the case of the karate network, there was a fixed probability p of transmission between any two individual that had a link in the network. For the other two networks, the probability of transmission was a function of the number of interactions in the original network. Given a fixed probability p , probability that an infectious node u infects node v was set to be $1 - (1 - p)^{k(u,v)}$, where $k(u, v)$ is number of 20 second interactions between u and v in the original dataset. This can be viewed as the probability of disease transmission given $k(u, v)$ interactions, with each interaction having an independent probability p of transmission. To keep confounding factors out of the model we assumed one possible transmission event with the above probabilities. I.e. we did not take into account possible infectious periods.

For the tests, we chose probabilities of transmission of $p \in (0.5, 0.95)$ for Karate network, $(0.0002, 0.002)$ for the Kenya Village network and $(0.002, 0.02)$ for the high school network. These probabilities were chosen to obtain a range from virtually no infected to almost the whole network being infected without intervention. To solve SAAIP, we generated 1000 scenarios for the karate network, 300 for the village network, and 100 for the High school network. These were chosen to have the most complex models solvable in reasonable time (less than an hour using Gurobi 8.1 on Ryzen 5 1600x 3.6 Ghz, 16GB of memory, Windows 10). We tested the immunization budgets from 1 to 10 for each network. Once the optimal strategies were obtained using every method, we ran 100000 simulations to get the expected number of infected within two significant digits. Full results are in the Appendix.

Overall, SAAIP intervention is always at least as good as the other interventions (with one caveat which we discuss in a bit). In some cases, e.g. Karate with transmission probability 0.95 and a budget of 8, it results in 30% less infected than the next best alternative. There is no dominant intervention strategy among the three we compared to SAAIP. For instance, depending on the intervention budget and probability of infection, intervention based on betweenness centrality can be the best (see Figure 5.4) or the worst (see Figure 5.5). Also, for some parameter values, SAAIP strictly dominates other interventions (e.g. see Figure 5.6 and 5.7). For other values, the intervention effectiveness can be quite close, e.g. Figure 5.8, where SAAIP pulls slightly ahead only for large values of p . Finally, in Figure 5.9, SAAIP-based intervention does marginally worse than the one based on optimizing R_0 , for small values of p . This is an artifact of only running 100 scenarios. After running more scenarios, SAAIP and R_0 settled on the same intervention. Also, in Figure 5.9 we can see that SAAIP outperforms everything for moderate values of p , but only performs equally well as the other interventions for high values of p . This again could be an artifact of the small number of samples. However, in this case, it is difficult to determine as the problems with the highest probability of infection are empirically the most

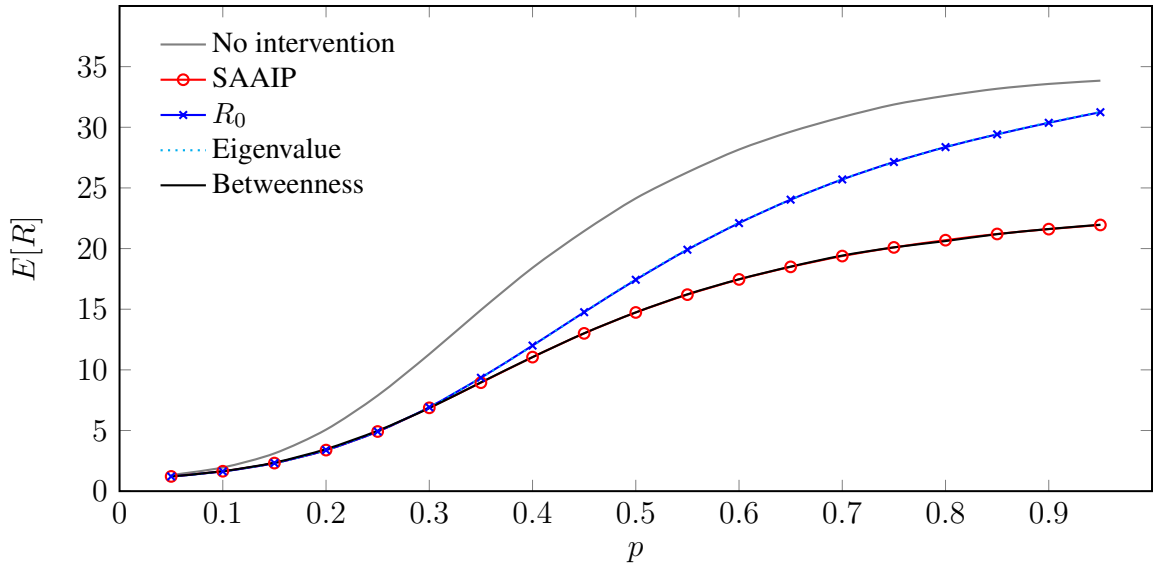


Figure 5.4: **Karate club**, $b = 1$. Expected infections $E[R]$ against probability of transmission p per interaction for the Karate network with intervention budget of 1. The curves (from top to bottom) correspond expected infections coming from intervention policy based on: no intervention, SAAIP, minimizing R_0 , removing node with highest betweenness centrality, and removing node with highest eigenvalue centrality. Note that in this case, R_0 and eigenvalue centrality based interventions produce the same policy. However, the optimal policy is the one based on SAAIP and betweenness centrality interventions.

difficult to solve (they result in larger scenarios).

Overall, SAAIP provides a good empirical solution for small to medium networks. It is also the only one that converges to the truly optimal intervention. Furthermore, the difference between the methods is most pronounced in the most dangerous cases of high probability of infection. It is especially important to find good interventions for these cases when the budgets are small. On the other hand, when the budget is large and probability of infection is small, pretty much any intervention will perform adequately. Furthermore, this pattern of differences in quality holds for three very different networks. One might have expected that contacts would be very well-mixed in the school social network, but according to the data they are not. The assumption of homogeneous mixing, which would make R_0 a true threshold and a good intervention measure, is again violated.

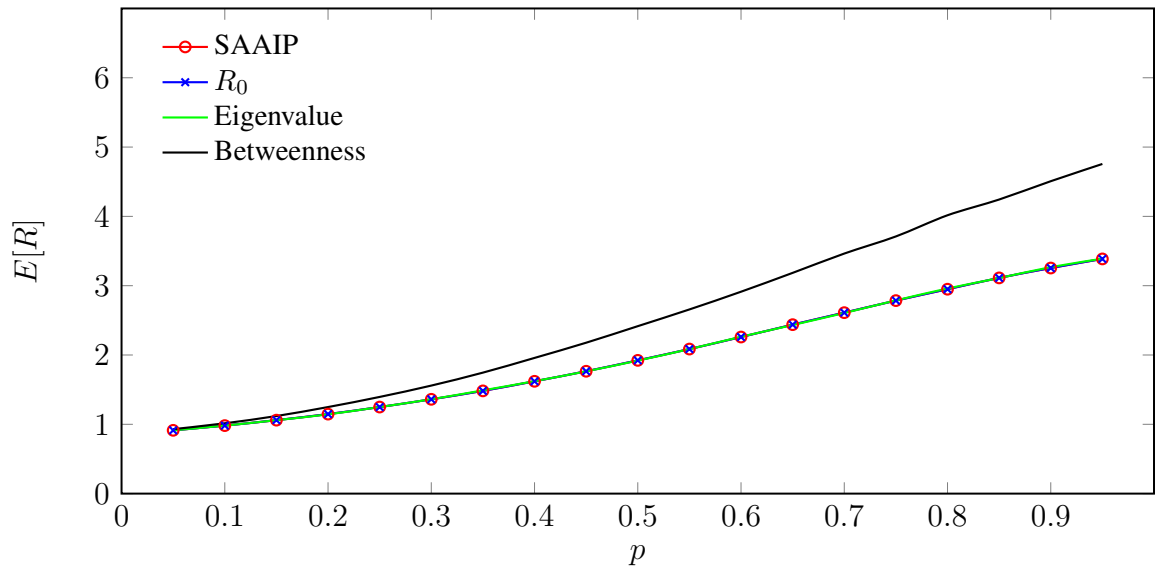


Figure 5.5: **Karate club**, $b = 5$. All interventions except the betweenness centrality produce the same policies.

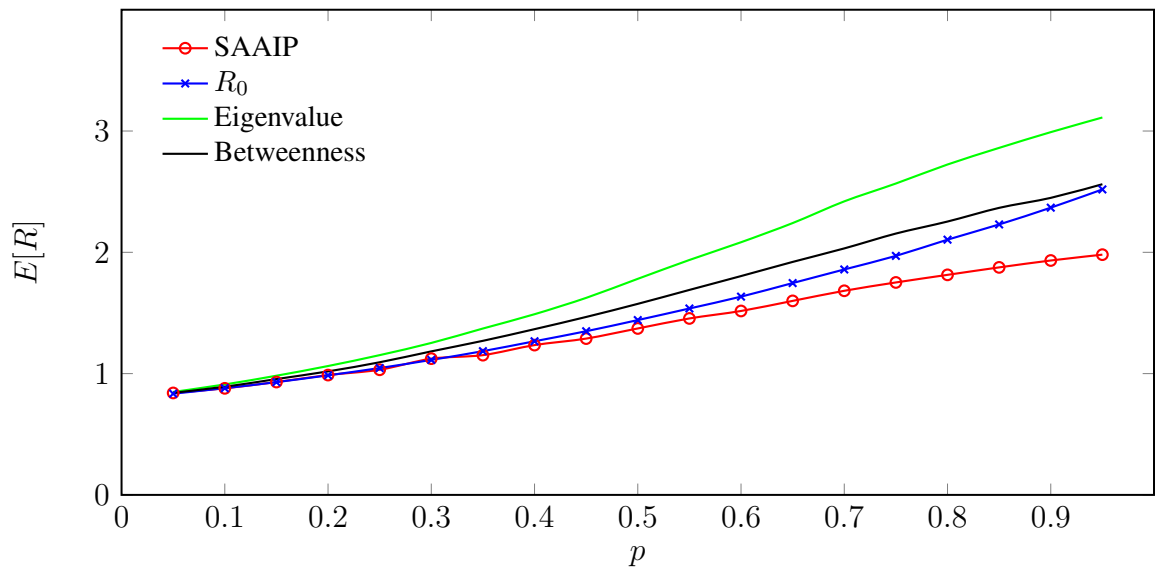


Figure 5.6: **Karate club**, $b = 7$. SAAIP based intervention strictly dominates all other interventions.

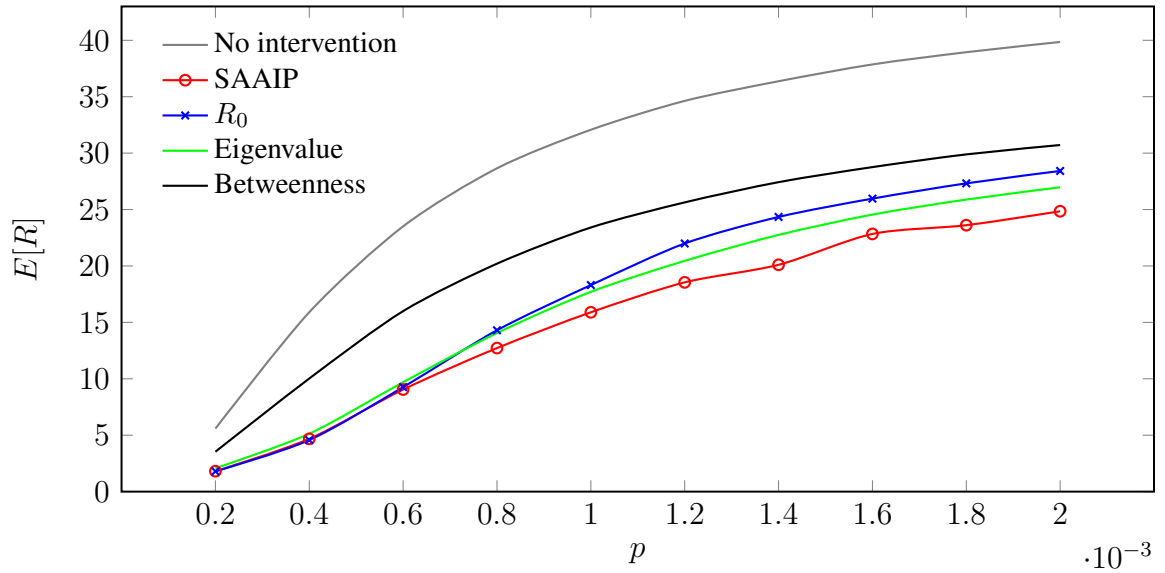


Figure 5.7: **Kenyan village**, $b = 5$. Again, SAAIP based intervention strictly dominates all other interventions.

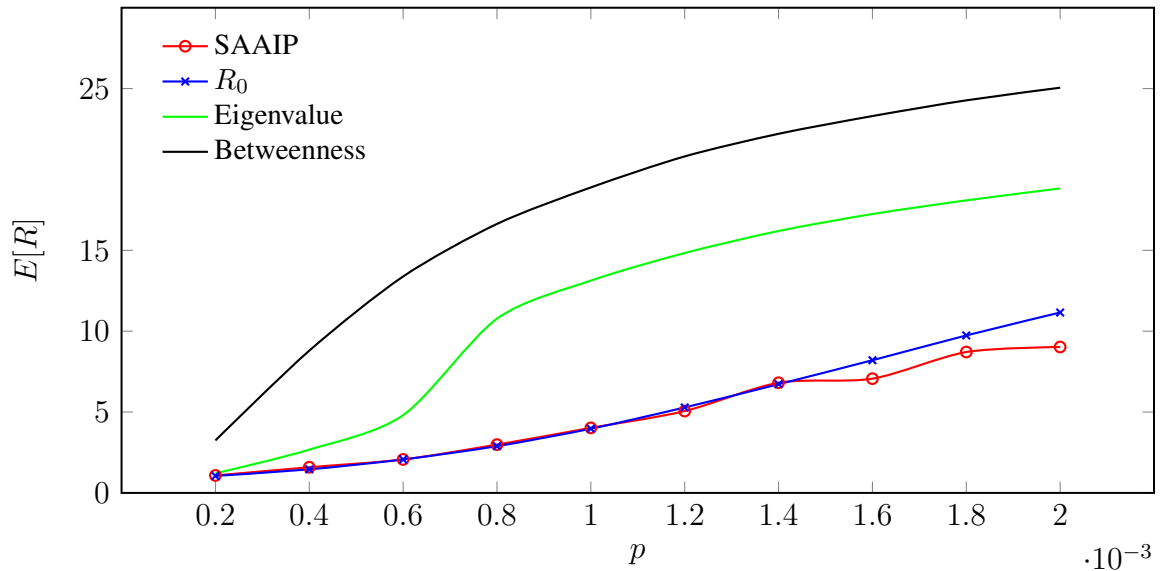


Figure 5.8: **Kenyan village**, $b = 10$. SAAIP based intervention policy dominates all other policies. However, it becomes different from R_0 based policy for large values of p

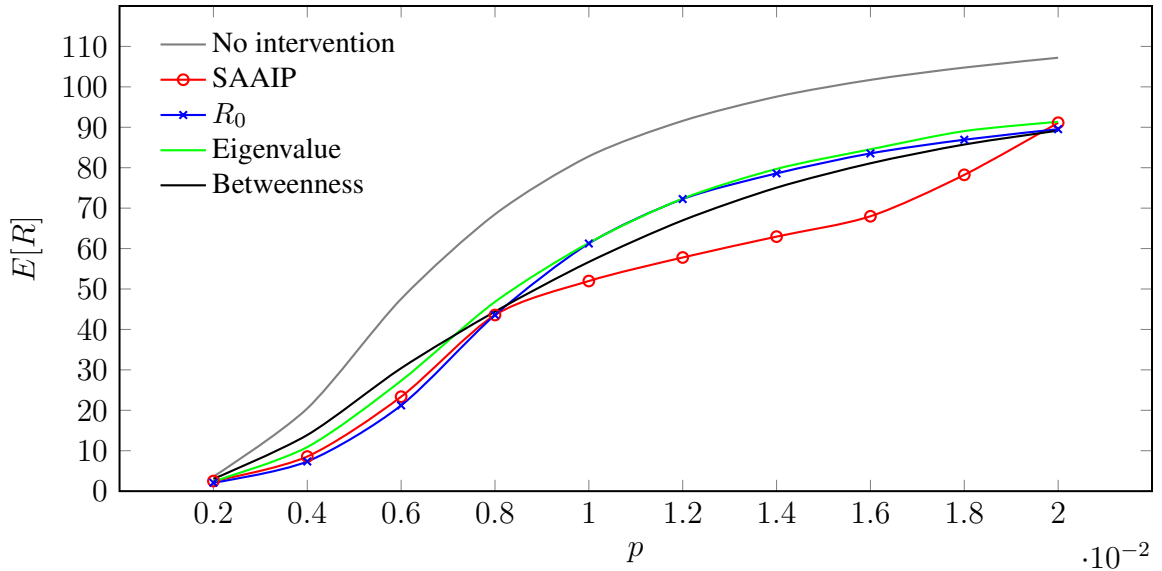


Figure 5.9: **French high school**, $b = 8$. R_0 based intervention slightly outperforms SAAIP for small values of p . Running SAAIP with larger number of scenarios fixes the issue. Also, notice that for very high values of p all policies perform about the same, but SAAIP is much better for moderate to high probabilities of transmission.

5.5 Conclusions

In this chapter, we discussed how to approach the problem of minimizing spread of infection in a network. We found evidence that obtaining or approximating optimal intervention is computationally hard. However, in practice, one can approximate and quantify the error for small networks. Furthermore, it performs considerably better than the leading alternatives described in the literature.

In terms of future research there is a vast number of interesting practical and analytical questions. We haven't been able to produce a strict result on hardness of approximation. Based on attempts at similar problems [refs] this seems to be difficult in general. Nevertheless, it would be an interesting avenue to pursue given the intricate combinatorial structure of the problem.

On the more practical side, there are a number of heuristic approaches one can pursue to solve SAAIP. An important avenue of research would be to characterize how far we can scale this approximation. I.e. it would be a good to know the size of the problem for

which we could obtain a reasonable (perhaps suboptimal) solution in practice. It would also be interesting to compare SAAIP based solutions to other proxies, perhaps appropriately modified shattering approaches, etc. Also it would be interesting to test this approach on different classes of networks which might be more prevalent in practice, for instance transportation, lower resolution population, or utility networks.

Appendices

APPENDIX A
THE LP FORMULATION

For completeness here we state an LP formulation functionally equivalent to that of [15].

Index sets: alternatives $i = 1, \dots, k$; states \mathbf{w} .

Decision variables: $q_{\mathbf{w}}$, the (unconditional) probability of (arriving and) terminating at state \mathbf{w} .

Auxiliary variables: $a_{\mathbf{w}}$, the probability of arriving at state \mathbf{w} .

Data:

$\mathbf{p} = (p_1, \dots, p_k)$, the probability vector; U , upper bound on # of trials; ζ , minimum PCS

Auxiliary data:

$t([\mathbf{w}]) =$ the # of components of the vector $[\mathbf{w}]$ equal to $[\mathbf{w}]_k$ (the # of alternatives tied for the most wins).

Defining Constraints of Auxiliary variables:

$$a_{\mathbf{0}} = 1;$$

$$a_{\mathbf{w}} = \sum_{i=1}^k (a_{\mathbf{w}-\mathbf{e}_i} - q_{\mathbf{w}-\mathbf{e}_i}) p_i \text{ (for all } \mathbf{w} \text{ except } \mathbf{w} = \mathbf{0})$$

Constraints:

$$0 \leq q_{\mathbf{w}} \leq a_{\mathbf{w}} \forall \mathbf{w}.$$

$$a_{\mathbf{w}} = q_{\mathbf{w}} \forall |\mathbf{w}| = U \text{ (upper bound on trials).}$$

$$\forall \mathbf{w} : \frac{q_{\mathbf{w}}}{\prod_{i=1}^k p_i^{w_i}} = \frac{q_{[\mathbf{w}]}}{\prod_{i=1}^k p_i^{[\mathbf{w}]_i}} \text{ (neutrality).}$$

$$\sum_{[\mathbf{w}]} q_{[\mathbf{w}]} / t([\mathbf{w}]) \geq \zeta \text{ (required PCS).}$$

Objective Function:

Minimize $\sum_{\mathbf{w}} |\mathbf{w}| q_{\mathbf{w}}$. (Minimize the expected number of trials upon termination).

APPENDIX B

COUNTEREXAMPLE TO A STRONGER STATEMENT THAN LEMMA 2.3.1

The policy with the termination rule: sample until one of the alternatives wins 6 trials or states $\{1, 4\}$ are reached (see Figure B.1) is a counterexample to Lemma 2.3.1 for sane, curtailed and bounded Markovian policies if we are confined to deterministic policies. Tables B.1 and B.2 show all of the deterministic policies that run up to 11 trials with their respective PCS and expected number of trials at $\mathbf{p} = (0.45, 0.55)$. The policies are sorted in ascending PCS. One can see that the holey policy has no deterministic unholey policy strictly dominating it. In other words, no unholey deterministic policy has at least as large PCS and a lesser or equal expected number of trials with at least one inequality being strict.

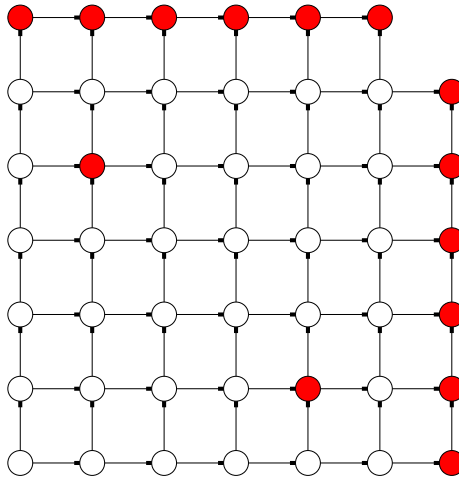


Figure B.1: **Holey policy**

Table B.1: PCS and expected number of trials for deterministic policies (part 1)

Policy States	PCS	Expected trials
(1, 0), (1, 1)	0.55	1
(2, 0), (2, 1)	0.57475	2.495
(2, 0), (3, 2), (3, 1)	0.58700125	3.235025
(2, 0), (3, 1), (4, 3), (4, 2)	0.593065619	3.601337375
(3, 0), (3, 2), (3, 1)	0.593126875	4.1100375
(2, 0), (5, 4), (3, 1), (5, 3), (4, 2)	0.596067481	3.782662001
(2, 0), (4, 1), (4, 3), (4, 2)	0.596097803	4.034468562
(6, 4), (3, 1), (2, 0), (4, 2), (6, 5), (5, 3)	0.597553403	3.87241769
(2, 0), (5, 4), (3, 1), (5, 2), (5, 3)	0.597568413	3.997061938
(6, 4), (3, 1), (6, 3), (2, 0), (4, 2), (6, 5)	0.598296364	3.97854566
(6, 4), (3, 1), (2, 0), (5, 2), (6, 5), (5, 3)	0.599797295	4.131695473
(2, 0), (5, 4), (5, 3), (4, 1), (4, 2)	0.600600597	4.306455501
(6, 4), (3, 1), (6, 3), (2, 0), (5, 2), (6, 5)	0.600911737	4.290887427
(6, 4), (3, 1), (6, 3), (2, 0), (6, 2), (6, 5)	0.601283217	4.407045349
(3, 0), (4, 2), (3, 1), (4, 3)	0.602223428	4.659506063
(6, 4), (2, 0), (4, 2), (4, 1), (6, 5), (5, 3)	0.60282948	4.441089035
(2, 0), (5, 4), (4, 1), (5, 2), (5, 3)	0.602851994	4.628055408
(2, 0), (5, 4), (5, 3), (5, 2), (5, 1)	0.603602459	4.862717877
(6, 4), (6, 3), (2, 0), (4, 2), (4, 1), (6, 5)	0.603943921	4.600280989
(6, 4), (5, 2), (2, 0), (4, 1), (6, 5), (5, 3)	0.606195318	4.830005709
(3, 0), (4, 2), (5, 4), (3, 1), (5, 3)	0.606726222	4.931493001
(3, 0), (4, 2), (4, 1), (4, 3)	0.606771705	5.309202844
(6, 4), (2, 0), (5, 1), (5, 2), (6, 5), (5, 3)	0.607317264	5.087107101
(6, 4), (5, 2), (6, 3), (2, 0), (4, 1), (6, 5)	0.60786698	5.06879364
(4, 2), (4, 1), (4, 3), (4, 0)	0.608287797	5.783268438
(6, 4), (6, 3), (2, 0), (6, 2), (4, 1), (6, 5)	0.608424201	5.243030523
(6, 4), (3, 0), (3, 1), (4, 2), (6, 5), (5, 3)	0.608955105	5.066126535
(3, 0), (5, 4), (3, 1), (5, 2), (5, 3)	0.608977619	5.253092908
(6, 4), (6, 3), (2, 0), (5, 1), (5, 2), (6, 5)	0.609174667	5.352427024
(6, 4), (6, 3), (2, 0), (6, 2), (5, 1), (6, 5)	0.609917628	5.584742868
(6, 4), (3, 0), (3, 1), (6, 3), (4, 2), (6, 5)	0.610069546	5.225318489
(6, 4), (6, 1), (6, 3), (2, 0), (6, 2), (6, 5)	0.610103368	5.708415517
(6, 4), (3, 0), (3, 1), (5, 2), (6, 5), (5, 3)	0.612320943	5.455043209

Table B.2: PCS and expected number of trials for deterministic policies (part 2)

Policy States	PCS	Expected trials
(3, 0), (4, 2), (5, 4), (4, 1), (5, 3)	0.613525895	5.717183251
(6, 4), (3, 0), (3, 1), (6, 3), (5, 2), (6, 5)	0.613992605	5.69383114
(6, 4), (3, 0), (3, 1), (6, 3), (6, 2), (6, 5)	0.614549826	5.868068023
(4, 2), (5, 4), (4, 0), (4, 1), (5, 3)	0.615792453	6.236580002
(6, 4), (3, 0), (4, 2), (4, 1), (6, 5), (5, 3)	0.61686922	5.919133553
(3, 0), (5, 2), (5, 4), (4, 1), (5, 3)	0.616902991	6.199583111
(3, 0), (5, 1), (5, 4), (5, 2), (5, 3)	0.618028689	6.551576815
(6, 4), (3, 0), (6, 3), (4, 2), (4, 1), (6, 5)	0.618540882	6.157921484
(6, 4), (4, 2), (4, 0), (4, 1), (6, 5), (5, 3)	0.619507258	6.460969226
(5, 2), (5, 4), (4, 0), (4, 1), (5, 3)	0.619544781	6.772579846
(5, 1), (5, 4), (4, 0), (5, 2), (5, 3)	0.621045713	7.241904784
(6, 4), (6, 3), (4, 2), (4, 1), (6, 5), (4, 0)	0.621364661	6.726289149
(5, 1), (5, 4), (5, 2), (5, 0), (5, 3)	0.621420945	7.491748518
(6, 4), (3, 0), (5, 2), (4, 1), (6, 5), (5, 3)	0.621917977	6.502508564
(6, 4), (3, 0), (5, 1), (5, 2), (6, 5), (5, 3)	0.623600897	6.888160651
(6, 4), (3, 0), (5, 2), (6, 3), (4, 1), (6, 5)	0.624425471	6.86069046
(6, 4), (5, 2), (4, 0), (4, 1), (6, 5), (5, 3)	0.625116989	7.109163682
(6, 4), (3, 0), (6, 3), (6, 2), (4, 1), (6, 5)	0.625261302	7.122045785
(6, 4), (3, 0), (6, 3), (5, 1), (5, 2), (6, 5)	0.626387	7.286140536
(6, 4), (5, 1), (4, 0), (5, 2), (6, 5), (5, 3)	0.627360881	7.623366465
(6, 4), (3, 0), (6, 3), (6, 2), (5, 1), (6, 5)	0.627501442	7.634614302
(6, 4), (3, 0), (6, 1), (6, 3), (6, 2), (6, 5)	0.627780052	7.820123275
(6, 4), (5, 2), (6, 3), (4, 1), (6, 5), (4, 0)	0.627903092	7.507143567
(6, 4), (5, 0), (5, 1), (5, 2), (6, 5), (5, 3)	0.627921854	7.884429661
(6, 4), (6, 3), (6, 2), (4, 1), (6, 5), (4, 0)	0.628831794	7.797538372
Holey Counterexample	0.628924664	8.060668447
(6, 4), (6, 3), (5, 1), (5, 2), (6, 5), (4, 0)	0.630518465	8.074410334
(6, 4), (6, 3), (5, 0), (5, 1), (5, 2), (6, 5)	0.631172308	8.348739526
(6, 4), (6, 3), (6, 2), (5, 1), (6, 5), (4, 0)	0.631818647	8.480963062
(6, 4), (6, 1), (6, 3), (6, 2), (6, 5), (4, 0)	0.632190127	8.728308359
(6, 4), (6, 3), (6, 2), (5, 0), (5, 1), (6, 5)	0.63256536	8.784331734
(6, 4), (6, 1), (6, 3), (6, 2), (5, 0), (6, 5)	0.633029711	9.093513355
(6, 4), (6, 1), (6, 0), (6, 3), (6, 2), (6, 5)	0.633122581	9.22413093

APPENDIX C
PROOFS OF THEOREMS IN CHAPTER 3

Proof of Theorem 3.2.1. Run k simulations. Let $R_i(\infty)$ be the number of removed nodes at the end of the i th simulation run. Let $\hat{R} = \frac{1}{k} \sum_{i=1}^k \frac{R_i(\infty)}{n}$, be the sample average over the k samples. Note that $0 \leq \frac{R_i(\infty)}{n} \leq 1$ for all i . For all $\epsilon > 0$, Hoeffding's Inequality [126] implies:

$$\begin{aligned} P\left(\left|\hat{R} - E(F)\right| > \epsilon\right) &= P\left(\left|\sum_{i=1}^k \frac{R_i(\infty)}{n} - kE(F)\right| > \epsilon k\right) \\ &\leq 2 \exp(-2\epsilon^2 k) \end{aligned}$$

Therefore, to get the required probability we need

$$k \geq \frac{1}{2\epsilon^2} \log \frac{2}{\delta}$$

samples. For a guarantee on relative error, *i.e.*,

$$P\left(\left|\hat{R} - E[F]\right| > \epsilon E[F]\right) < \delta$$

a similar calculation shows that we need

$$k \geq \frac{1}{2\epsilon^2 E[F]^2} \log \frac{2}{\delta}.$$

As $E[F]^2 \geq 1/|V|^2$ (we start with at least one infected node), we need

$$k \geq \frac{|V|^2}{2\epsilon^2} \log \frac{2}{\delta}$$

samples to get an estimate with the required accuracy.

To generate one sample we can perform a search in the graph, following only the edges that have transmitted the infection. Therefore, it takes $O(|E|)$ time to run each simulation. The $O(|E|)$ time bound per simulation applies to any epidemic model for which the number of node states and the number of time steps a node may remain in a transient state are bounded. In the simple epidemic model of the theoretical analysis, there are three nodes states (S, I, R) and a node may remain in the transient state I for one time step. In the case study, the number of states and duration lengths are larger but still bounded. \square

Proof of Theorem 3.2.2. The basic reproduction number of H is

$$R_0^H = \frac{(|E| + 2)p}{|V|} > \frac{|E|p}{|V|} = R_0,$$

For the expected fraction of infected we need to choose the connecting vertex appropriately.

Let A_v be an event that the epidemic started from vertex v . Then

$$E[F] = \sum_{v \in V} E[F|A_v]P(A_v) = \frac{1}{|V|} \sum_{v \in V} E[F|A_v]$$

There exists u such that $E[F|A_u] \leq E[F]$. Choose this u to connect the components in H . Label them u_1 and u_2 . Let B be an event that there was a transmission along the edge (u_1, u_2) (irrespective of whether u_1 or u_2 actually got infected). Then, the expected fraction of infected of H is

$$\begin{aligned} E[F^H] &= E[F^H|B]p + E[F^H|B^c](1-p) \\ &\leq \frac{1}{2}p(E[F] + E[F|A_u]) + \frac{1}{2}(1-p)E[F] \\ &\leq \frac{1}{2}(1+p)E[F] < E[F] \end{aligned}$$

\square

Proof of Theorem 3.2.3. The basic reproduction number of H is

$$R_0^H = \frac{(|E| + 2m - 2)p}{|V|} > \frac{|E|p}{|V|} = R_0, \quad (\text{C.1})$$

Equation (C.1) also implies that $R_0^H \rightarrow \infty$ as $m \rightarrow \infty$.

As in Theorem 3.2.2, there exists u such that $E[F|A_u] \leq E[F]$. Choose this u to connect the components in H . That is, label the copies of G from 1 to m and vertices corresponding to v as v_s . Vertex u_s is connected to u_{s+1} for each $s \in \{1, \dots, m-1\}$. Then

$$\begin{aligned} E[F^H] &\leq \frac{1}{m} \sum_{s=0}^{m-1} (1+s)^2 p^s E[F] \\ &< \frac{1+p}{m(1-p)} E[F] \end{aligned}$$

Let $\rho = |E|/|V|$ be the edge density of G . Then the edge density of H is $(m|E| + m - 1)/m|V| = \rho + \frac{1-1/m}{|V|} \geq \rho$. □

Proof of Theorem 3.2.4. Let H_n be a family of (b, d) expander graphs with n vertices, where d is the degree of each vertex (can fix say $d = 4$ so that they exists for all n large enough). They exist, as for a fixed $d \geq 3$ there exists a $b > 0$ s.t. a random (n, d) -graph, i.e. a d -regular graph on n vertices, is a b -expander with probability tending to 1 as n goes to infinity [127].

Fix $c = 1/2$. Then by Theorem 1.3 of [128] there exist $0 < q_1(d) < q_2(c) < 1$ and $p^* \in [q_1, q_2]$ such that for all n large enough, $H_n(p^*)$, a graph obtained from H_n by independently removing edges with probability $1 - p^*$, has a component of size at least $n/2$ with high probability. In particular, for n large enough it has one with probability at least $1/2$.

Following [129], tracing an epidemic on H_n from vertex v is equivalent to performing a breadth first search on $H_n(p^*)$ from v and labeling visited nodes as removed. In other words, the expected epidemic size with infection probability p^* on H_n equals the expected

size of the connected component reachable from a randomly picked vertex in $H_n(p^*)$.

Now, as H_n s are regular graphs, their basic reproduction numbers are dp^* . Let

$$\alpha = \max\left(\frac{dp^*}{R_0} - 1, 0\right)$$

Construct G_n s by adding $\lceil \alpha n \rceil$ singletons to H_n s. This way, epidemics with transmission probability p^* running on G_n s have basic reproduction numbers $R_0^n \leq R_0$.

Let A_n be the event that $H_n(p^*)$ has component of size at least $n/2$ and epidemic starts in this component. By construction, $P(A_n) \geq 1/4$. Therefore, the expected size of the epidemic on G_n is

$$\begin{aligned} E[R_G^n(\infty)] &= \alpha + (1 - \alpha)E[R_H^n(\infty)] \\ &\geq \alpha + (1 - \alpha)P(A_n) \frac{n}{2} \geq \alpha + (1 - \alpha) \frac{n}{8} \end{aligned}$$

Set $F = (1 - \alpha)/8$. Then $E[F_n] \geq F$. □

APPENDIX D
EXPERIMENTAL RESULTS FOR THE OPTIMAL INTERVENTION

Table D.1: **Karate network intervention results.** No intervention, SAIP, and R_0 based interventions

Budget	Probability of Transmission									
	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
	No Intervention									
1	1.32964	3.1123	7.88219	14.93763	21.44262	26.28637	29.63256	31.87657	33.17729	33.84301
	SAIP									
1	1.21527	2.3147	4.917434	8.942993	13.00822	16.201863	18.491325	20.095647	21.203024	21.946116
2	1.145674	1.679724	2.819578	4.806952	7.599535	10.766899	13.754431	16.130175	17.051194	17.578835
3	1.028443	1.387792	1.991844	2.908326	4.172068	5.756109	7.549538	9.312652	10.920558	12.190297
4	0.966919	1.203204	1.500378	1.90676	2.408921	2.993559	3.627137	4.290248	4.922117	5.493891
5	0.91311	1.05985	1.248891	1.482138	1.765384	2.085694	2.437691	2.784719	3.112255	3.386155
6	0.873706	0.991094	1.145728	1.316967	1.486838	1.717664	1.891966	2.089341	2.265856	2.410718
7	0.840498	0.930525	1.031577	1.152061	1.288697	1.454763	1.599544	1.751555	1.87576	1.980661
8	0.809366	0.886942	0.944947	1.037958	1.107552	1.204668	1.311927	1.488452	1.554846	1.711712
9	0.756341	0.801316	0.850593	0.902763	0.993865	1.025852	1.094911	1.168682	1.248874	1.333185
10	0.726798	0.773588	0.784476	0.816541	0.850626	0.885288	0.920743	0.959558	0.999862	1.035768
	R_0									
1	1.210327	2.3147	4.917434	9.348537	14.755162	19.905834	24.031943	27.13622	29.417679	31.242658
2	1.102759	1.679724	2.819578	4.806952	7.599535	10.766899	13.754431	16.183985	18.134774	19.835033
3	1.028443	1.387792	1.991844	2.908326	4.172068	5.756109	7.549538	9.312652	10.920558	12.190297
4	0.966241	1.203204	1.559988	2.072672	2.797987	3.741642	4.87625	6.100689	7.231518	8.157796
5	0.91311	1.05985	1.248891	1.482138	1.765384	2.085694	2.437691	2.784719	3.112255	3.386155
6	0.872293	0.990105	1.131276	1.297544	1.493744	1.717664	1.963631	2.235308	2.536109	2.874246
7	0.834477	0.929661	1.045876	1.185139	1.349405	1.536683	1.746477	1.970503	2.229864	2.518418
8	0.79555	0.862038	0.944947	1.037509	1.150196	1.287657	1.449486	1.653506	1.909418	2.22534
9	0.756544	0.801547	0.850593	0.903511	0.962925	1.025282	1.09376	1.168682	1.248874	1.336266
10	0.721227	0.752453	0.788507	0.826377	0.867155	0.913623	0.96328	1.016631	1.076171	1.142294

Table D.2: **Karate network intervention results continued.** Interventions based on betweenness and eigenvector centralities.

Budget	Probability of Transmission									
	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
Betweenness										
1	1.21343	2.35298	4.98246	8.96856	13.0169	16.22994	18.51281	20.09906	21.18833	21.96591
2	1.10322	1.68153	2.82785	4.79449	7.59386	10.76396	13.74971	16.13725	18.12571	19.84504
3	1.02735	1.38486	1.98105	2.89054	4.17167	5.73257	7.53048	9.31734	10.92053	12.18943
4	0.97014	1.18881	1.50963	1.91172	2.41427	2.9907	3.62633	4.28066	4.93286	5.49148
5	0.92986	1.11931	1.39496	1.74561	2.17735	2.65715	3.18495	3.71057	4.24314	4.75627
6	0.89648	1.08567	1.33739	1.66604	2.05951	2.50169	2.94831	3.43598	3.86683	4.23673
7	0.84147	0.95494	1.09349	1.27031	1.46707	1.68997	1.92081	2.15471	2.36654	2.56143
8	0.80997	0.913	1.04088	1.19716	1.38596	1.5841	1.79888	2.00588	2.19202	2.35522
9	0.78109	0.88674	1.01686	1.17652	1.35739	1.55794	1.77978	1.98228	2.17069	2.32933
10	0.74088	0.81476	0.91423	1.02873	1.16108	1.30823	1.48088	1.65432	1.86479	2.03007
Eigenvector										
1	1.2125	2.30306	4.93248	9.3387	14.7632	19.93775	24.03008	27.10994	29.42662	31.25929
2	1.10409	1.67668	2.81295	4.77483	7.59833	10.72626	13.72843	16.1872	18.13981	19.82256
3	1.03473	1.41418	2.01916	2.97896	4.40477	6.28983	8.6737	11.322	14.04014	16.54693
4	0.9666	1.19493	1.50186	1.91073	2.40311	2.98666	3.63637	4.29421	4.92859	5.51837
5	0.91302	1.0592	1.24905	1.48987	1.76398	2.08491	2.43232	2.78746	3.10995	3.38912
6	0.88086	1.02005	1.20798	1.4276	1.70938	2.02075	2.36943	2.70633	3.01937	3.31578
7	0.84862	0.9826	1.15232	1.37171	1.62438	1.93684	2.24108	2.56726	2.8611	3.11165
8	0.81354	0.93431	1.09177	1.2943	1.53061	1.81182	2.11652	2.41414	2.69771	2.91244
9	0.77084	0.86886	0.98103	1.12147	1.29261	1.47771	1.66728	1.85403	2.03159	2.16478
10	0.74477	0.83451	0.95256	1.09271	1.2584	1.44531	1.64517	1.83262	2.01686	2.13463

Table D.3: **Kenyan village network intervention results.** No intervention, SAAIP, and R_0 based interventions

Budget	Probability of Transmission									
	0.0002	0.0004	0.0006	0.0008	0.001	0.0012	0.0014	0.0016	0.0018	0.002
	No intervention									
5.58121	15.90117	23.50928	28.65466	32.06997	34.62833	36.37092	37.86056	38.94417	39.85113	
	SAAIP									
1	3.94916	12.83871	20.18595	25.28262	28.79096	30.65775	32.74353	34.7197	35.92959	36.60635
2	3.41487	9.84183	17.23662	22.51947	26.04812	27.41418	29.27145	30.77106	34.30853	34.50827
3	2.52426	7.58493	14.13353	18.95877	22.48716	25.51124	26.33116	29.19714	29.0227	29.89682
4	2.12416	5.77927	11.22632	16.32665	19.39202	22.10798	23.91596	25.13795	26.38513	26.89334
5	1.81281	4.67915	9.0476	12.72213	15.89298	18.54461	20.10366	22.83756	23.6117	24.84743
6	1.59642	3.30512	7.03108	10.42476	12.42406	14.92148	17.81847	18.88758	21.35513	21.79301
7	1.40403	2.84917	4.95122	7.86098	9.4665	11.17687	12.88155	14.77625	16.5659	19.46507
8	1.31294	2.17096	3.67085	6.28484	8.08203	9.00533	10.45482	12.1842	13.50726	14.73901
9	1.21697	1.72646	2.64056	4.3554	6.08086	7.00718	8.72097	9.8926	10.59536	12.00162
10	1.0783	1.58632	2.06807	2.98674	4.02128	5.05982	6.81338	7.06443	8.71118	9.02127
	R_0									
1	3.94916	12.83871	20.94006	26.21054	29.73531	32.27898	34.19755	35.59497	36.64125	37.57813
2	3.03778	9.84183	18.01587	23.80468	27.49371	29.8995	31.75366	33.1979	34.19624	35.1744
3	2.52069	7.62341	14.88763	20.52165	24.88803	27.45352	29.3674	30.77586	31.95163	32.91018
4	2.07996	5.77927	12.23237	17.86311	21.73432	24.52937	27.43876	28.94189	30.03352	31.04296
5	1.79784	4.56871	9.26109	14.30064	18.30441	21.992	24.34355	25.9675	27.31998	28.41953
6	1.54602	3.26828	6.75434	11.32164	15.60892	18.83066	21.18307	22.9502	24.42779	25.45956
7	1.38589	2.6516	5.01439	8.40592	12.04631	15.38901	17.83511	19.87671	21.53607	22.84012
8	1.25047	2.08596	3.64365	6.32946	8.59585	10.4753	12.7906	14.84236	16.712	18.20324
9	1.14485	1.75675	2.7378	4.10582	5.8247	7.73537	9.70579	11.56747	13.10752	14.58418
10	1.05222	1.45643	2.06807	2.89045	3.96895	5.28428	6.72108	8.20343	9.73578	11.15429

Table D.4: **Kenyan village network intervention results continued.** Interventions based on betweenness and eigenvector centralities.

Budget	Probability of Transmission									
	0.0002	0.0004	0.0006	0.0008	0.001	0.0012	0.0014	0.0016	0.0018	0.002
Betweenness										
1	5.4565	15.43964	22.62598	27.2932	30.55355	32.94989	34.63858	36.05548	37.01873	37.81623
2	5.46003	15.38153	22.51507	27.10469	30.14847	32.39955	34.13214	35.29728	36.42666	37.22914
3	4.99861	13.51196	19.90816	24.46383	27.81248	30.17566	31.99557	33.27963	34.37565	35.28373
4	4.66747	12.53691	18.47984	22.71869	25.75923	28.01589	29.83874	31.1814	32.20045	33.0745
5	3.5343	10.02414	15.99664	20.20189	23.40847	25.63392	27.42654	28.76101	29.88773	30.71704
6	3.51006	9.9427	15.71064	19.78127	22.73391	24.92304	26.58805	27.75908	28.81339	29.52803
7	3.45741	9.66411	15.36036	19.26288	22.1009	24.04039	25.63422	26.65423	27.55411	28.25827
8	3.44163	9.59865	15.1742	18.91929	21.64877	23.62478	24.94732	26.05884	26.8609	27.50603
9	3.25798	8.77936	13.41094	16.73913	18.97383	20.87728	22.38697	23.56951	24.56995	25.23644
10	3.24403	8.79609	13.3827	16.63234	18.88841	20.80577	22.20382	23.30108	24.27316	25.05401
Eigenvector										
1	3.95056	12.87632	21.03232	26.29836	29.85137	32.31214	34.16383	35.51025	36.6371	37.61947
2	3.34349	9.77739	17.98255	23.75216	27.36365	29.77338	31.76599	33.07404	34.2291	35.19775
3	3.15082	7.65175	14.82458	21.0198	24.90224	26.54675	29.42374	30.85842	31.87529	32.98658
4	2.4897	6.13706	11.95706	17.32383	21.26367	23.9224	25.98567	27.46561	28.69985	29.66379
5	2.07168	5.1315	9.69357	14.05633	17.70494	20.44955	22.75104	24.55385	25.88295	26.97308
6	1.6676	3.68445	8.64178	12.32095	15.39429	17.89042	19.88933	21.45038	22.97584	23.95709
7	1.56882	3.17662	8.40986	11.99326	14.92545	17.26351	19.04278	20.51361	21.80903	22.76549
8	1.45445	2.87908	7.97589	11.20031	13.70656	15.69379	17.30858	18.57194	19.64747	20.4842
9	1.41869	2.82607	5.92654	11.0099	13.37478	15.20754	16.69198	17.7492	18.78736	19.50673
10	1.20987	2.674	4.80539	10.77499	13.12714	14.83312	16.19571	17.2438	18.08741	18.82726

Table D.5: **French high school network intervention results.** No intervention, SAAP, and R_0 based interventions

Budget	Probability of Transmission									
	0.002	0.004	0.006	0.008	0.01	0.012	0.014	0.016	0.018	0.02
	No Intervention									
3.60079	20.46069	47.55023	68.48839	82.8141	91.5849	97.56978	101.72032	104.77148	107.20241	
	SAAP									
1	3.28264	18.11091	43.56916	63.11869	80.51864	87.29867	96.22923	100.12052	100.89735	103.05418
2	3.3228	15.2971	40.15111	58.88461	74.19742	82.74008	92.98916	97.33605	98.62877	100.89548
3	2.82949	12.88602	39.74621	54.01094	67.16337	77.51423	86.98587	90.67191	95.0487	98.15802
4	2.70366	11.68737	33.94259	54.27072	62.85641	74.01797	84.31386	92.57728	92.96282	94.4171
5	2.71988	9.60294	32.06329	53.60108	59.41426	67.75428	83.34615	88.06411	90.89038	91.45279
6	2.59217	9.84073	29.52307	46.33172	59.03966	65.22972	80.97035	80.81423	93.0784	92.19751
7	2.5913	8.10356	26.69553	47.61164	59.05299	59.80443	76.20948	73.271	87.66158	90.48395
8	2.50058	8.55631	23.3638	43.57197	51.97873	57.79775	62.95748	67.99194	78.23791	91.1242
9	2.24846	7.45697	22.11966	37.8729	54.41478	63.20169	59.68575	70.6535	85.35311	87.32679
10	2.51214	8.43192	21.40334	37.22387	50.70239	58.7415	59.86384	68.00773	79.78523	83.81899
	R_0									
1	3.24359	17.95573	43.73922	65.39953	80.10335	89.30765	95.44088	99.76598	102.67641	105.24226
2	2.92786	14.72938	40.88322	63.02335	78.06536	87.56769	93.48165	97.74721	100.59406	103.22582
3	2.79955	13.18204	38.06878	60.76896	75.53715	85.32597	91.36462	94.73685	97.92237	100.31174
4	2.56468	11.25432	34.93811	57.0813	72.28822	81.95608	88.57641	93.13753	96.24635	98.61639
5	2.40029	9.72102	31.25925	53.66324	69.16917	78.92167	86.13774	90.90985	93.83163	96.20794
6	2.30827	8.47438	27.82742	50.66584	66.46552	76.21937	83.08583	87.48348	91.10714	94.45359
7	2.2038	7.7836	24.65638	47.31783	63.77244	74.26193	80.88027	85.57327	88.72799	91.59327
8	2.09305	7.34612	21.21302	43.56132	61.24991	72.24702	78.59137	83.53803	86.92437	89.52925
9	1.97006	6.34881	19.1925	39.73843	58.48343	69.66292	76.76187	81.74555	85.25248	87.58453
10	1.89325	5.55998	17.25475	35.82947	54.58171	66.37321	73.91198	78.98437	82.49842	85.36793

Table D.6: **French high school network intervention results continued.** Interventions based on betweenness and eigenvector centralities.

Budget	Probability of Transmission									
	0.0002	0.0004	0.0006	0.0008	0.001	0.0012	0.0014	0.0016	0.0018	0.002
Betweenness										
1	3.54747	19.2719	43.47666	63.21217	77.97506	87.71709	94.39333	99.11482	102.07704	104.67355
2	3.50386	18.56819	40.85622	58.95313	73.11604	83.48473	90.59066	95.67413	99.35	101.95268
3	3.27791	16.47442	37.44896	55.66023	69.97889	80.73636	88.03643	92.93066	96.40679	99.3493
4	3.28441	16.39941	37.14989	55.18746	69.6693	79.94346	86.9772	91.93763	95.67821	98.19441
5	3.22242	15.71692	34.51061	50.80513	64.38382	75.26437	83.28488	88.91728	93.26358	95.99447
6	3.15991	15.07312	33.24933	48.67076	61.99705	72.76062	80.80133	86.66491	90.62913	93.59339
7	3.11782	14.45375	31.6132	45.7846	57.95806	68.53817	76.81241	82.72648	87.37178	90.89481
8	3.04039	13.8887	30.40945	44.41307	56.69088	66.97929	75.06604	81.09414	85.74082	89.17436
9	2.79132	12.39262	28.27466	42.30865	54.56449	65.10556	72.92942	79.08878	83.55371	86.94351
10	2.64675	11.03238	25.67748	40.06488	52.17734	62.63105	70.77698	77.20925	81.89144	85.17282
Eigenvector										
1	3.31617	17.79058	44.88652	66.19439	80.73667	89.76243	95.77228	99.8154	102.81888	105.0947
2	3.185	16.19591	42.11978	63.77038	78.49521	87.66614	93.78128	97.76114	100.86416	103.20564
3	2.84724	13.93081	38.97391	61.0801	75.90105	85.34387	91.56618	95.64838	98.75582	101.45217
4	2.81073	13.21472	35.41868	57.51782	73.5693	82.92111	89.49566	93.8637	97.01992	99.36949
5	2.70203	12.67823	32.25191	54.57018	69.93765	80.77286	87.10399	91.75647	95.11405	97.1105
6	2.63061	11.53143	29.33927	51.62959	66.65647	78.53797	85.2256	89.88243	93.18309	95.5692
7	2.51503	11.09073	27.95223	49.0783	64.48287	75.02363	82.70164	87.53903	90.96915	93.24889
8	2.44168	10.9052	27.30954	46.82339	61.40219	72.35004	79.70327	84.54524	89.0588	91.34744
9	2.41417	9.19021	24.44808	42.84812	58.87221	70.11535	77.38783	82.02616	86.69558	89.51737
10	2.36986	8.32424	22.54992	41.51893	56.69255	67.60195	75.07514	80.00645	83.57759	87.29131

REFERENCES

- [1] C. for Medicare, M. Services, *et al.*, *National health expenditures 2018 highlights*, 2018.
- [2] P. Eveborn, M. Rönqvist, H. Einarsdóttir, M. Eklund, K. Lidén, and M. Almroth, “Operations research improves quality and efficiency in home care,” *Interfaces*, vol. 39, no. 1, pp. 18–34, 2009.
- [3] A. Rais and A. Viana, “Operations research in healthcare: A survey,” *International transactions in operational research*, vol. 18, no. 1, pp. 1–31, 2011.
- [4] J. D. Gibbons, I. Olkin, and M. Sobel, “An introduction to ranking and selection,” *The American Statistician*, vol. 33, no. 4, pp. 185–195, 1979.
- [5] S.-H. Kim and B. L. Nelson, “Recent advances in ranking and selection,” in *2007 Winter Simulation Conference*, IEEE, 2007, pp. 162–172.
- [6] T. J. Moore, H. Zhang, G. Anderson, and G. C. Alexander, “Estimated costs of pivotal trials for novel therapeutic agents approved by the us food and drug administration, 2015-2016,” *JAMA internal medicine*, vol. 178, no. 11, pp. 1451–1457, 2018.
- [7] D. L. Bhatt and C. Mehta, “Adaptive designs for clinical trials,” *New England Journal of Medicine*, vol. 375, no. 1, pp. 65–74, 2016.
- [8] D. A. Berry and C.-H. Ho, “One-sided sequential stopping boundaries for clinical trials: A decision-theoretic approach,” *Biometrics*, pp. 219–227, 1988.
- [9] A. Durand, C. Achilleos, D. Iacovides, K. Strati, G. D. Mitsis, and J. Pineau, “Contextual bandits for adapting treatment in a mouse model of de novo carcinogenesis,” in *Machine Learning for Healthcare Conference*, 2018, pp. 67–82.
- [10] N. T. Rindtorff, M. Lu, N. A. Patel, H. Zheng, and A. D’Amour, “A biologically plausible benchmark for contextual bandit algorithms in precision oncology using in vitro data,” *arXiv preprint arXiv:1911.04389*, 2019.
- [11] R. Bechhofer and S. Elmaghraby, “Morse (1959). a single-sample multiple-decision procedure for selecting the multinomial event which has the largest probability,” *Ann. Math. Statist.*, vol. 30, pp. 102–1,

- [12] M. Sobel and M. J. Huyett, “Selecting the best one of several binomial populations,” *Bell System Technical Journal*, vol. 36, no. 2, pp. 537–576, 1957.
- [13] T Cacoullos and M Sobel, “An inverse sampling procedure for selecting the most probable event in a multinomial distribution,” *Multivariate Analysis*, pp. 423–455, 1966.
- [14] J. T. Ramey and K. Alam, “A sequential procedure for selecting the most probable multinomial event,” *Biometrika*, vol. 66, no. 1, pp. 171–173, 1979.
- [15] E. Tollefson, D. Goldsman, A. Kleywegt, and C. Tovey, “Optimal selection of the most probable multinomial alternative,” *Sequential Analysis*, vol. 33, no. 4, pp. 491–508, 2014.
- [16] C. A. Tovey, “The slippage configuration is always the least favorable configuration for two alternatives,” *Sequential Analysis*, vol. 33, no. 4, pp. 509–518, 2014.
- [17] D. S. Jones, S. H. Podolsky, and J. A. Greene, “The burden of disease and the changing task of medicine,” *New England Journal of Medicine*, vol. 366, no. 25, pp. 2333–2338, 2012.
- [18] N. P. Johnson and J. Mueller, “Updating the accounts: Global mortality of the 1918–1920” spanish” influenza pandemic,” *Bulletin of the History of Medicine*, pp. 105–115, 2002.
- [19] D. Bernoulli, “Essai d’une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l’inoculation pour la prévenir,” *Histoire de l’Acad., Roy. Sci.(Paris) avec Mem*, pp. 1–45, 1760.
- [20] W. O. Kermack and A. G. McKendrick, “A contribution to the mathematical theory of epidemics,” *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, vol. 115, no. 772, pp. 700–721, 1927.
- [21] W. H. Frost, “Some conceptions of epidemics in general,” *American Journal of Epidemiology*, vol. 103, no. 2, pp. 141–151, 1976.
- [22] D. Schenzle, “Problems in drawing epidemiological inferences by fitting epidemic chain models to lumped data,” *Biometrics*, pp. 843–847, 1982.
- [23] J. M. Read, K. T. Eames, and W. J. Edmunds, “Dynamic social networks and the implications for the spread of infectious disease,” *Journal of The Royal Society Interface*, vol. 5, no. 26, pp. 1001–1007, 2008.

- [24] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, “Epidemic processes in complex networks,” *Reviews of modern physics*, vol. 87, no. 3, p. 925, 2015.
- [25] P. Flu, “A national framework for responding to an influenza pandemic,” *Cabinet Office*, 2007.
- [26] N. Gay, L. Hesketh, P Morgan-Capner, and E. Miller, “Interpretation of serological surveillance data for measles using mathematical models: Implications for vaccine strategy,” *Epidemiology & Infection*, vol. 115, no. 1, pp. 139–156, 1995.
- [27] W. H. Organization *et al.*, “Transmission dynamics and impact of pandemic influenza a (h1n1) 2009 virus= dynamique de transmission et effets du virus de la grippe pandémique a (h1n1) 2009,” *Wkly Epidemiol Rec*, vol. 84, no. 46, pp. 481–484, 2009.
- [28] P. Holme and N. Masuda, “The basic reproduction number as a predictor for epidemic outbreaks in temporal networks,” *PloS one*, vol. 10, no. 3, e0120567, 2015.
- [29] C. Nowzari, V. M. Preciado, and G. J. Pappas, “Analysis and control of epidemics: A survey of spreading processes on complex networks,” *IEEE Control Systems Magazine*, vol. 36, no. 1, pp. 26–46, 2016.
- [30] R. Pastor-Satorras and A. Vespignani, “Immunization of complex networks,” *Physical review E*, vol. 65, no. 3, p. 036 104, 2002.
- [31] L. K. Gallos, F. Liljeros, P. Argyrakis, A. Bunde, and S. Havlin, “Improving immunization strategies,” *Physical Review E*, vol. 75, no. 4, p. 045 104, 2007.
- [32] R. E. Bechhofer, “A single-sample multiple decision procedure for ranking means of normal populations with known variances,” *The Annals of Mathematical Statistics*, pp. 16–39, 1954.
- [33] R. E. Bechhofer, J. Kiefer, and M. Sobel, *Sequential identification and ranking procedures: with special reference to Koopman-Darmois populations*. University of Chicago Press, 1968, vol. 3.
- [34] K. Alam, “On selecting the most probable category,” *Technometrics*, vol. 13, no. 4, pp. 843–850, 1971.
- [35] J. D. Gibbons, I. Olkin, and M. Sobel, *Selecting and ordering populations: a new statistical methodology*. SIAM, 1999.

- [36] R. E. Bechhofer and R. V. Kulkarni, "Closed sequential procedures for selecting the multinomial events which have the largest probabilities," *Communications in Statistics-Theory and Methods*, vol. 13, no. 24, pp. 2997–3031, 1984.
- [37] P. Chen, "Truncated selection procedures for the most probable event and the least probable event," *Annals of the Institute of Statistical Mathematics*, vol. 44, no. 4, pp. 613–622, 1992.
- [38] R. E. Bechhofer and D. M. Goldsman, "Truncation of the bechhofer-kiefer-sobel sequential procedure for selecting the multinomial event which has the largest probability," *Communications in Statistics-Simulation and Computation*, vol. 14, no. 2, pp. 283–315, 1985.
- [39] D. M. Goldsman and R. E. Bechhofer, "Truncation of the bechhofer-kiefer-sobel sequential procedure for selecting the multinomial event which has the largest probability (ii): Extended tables and an improved procedure," *Communications in Statistics-Simulation and Computation*, vol. 15, no. 3, pp. 829–851, 1986.
- [40] P. Chen, "Closed inverse sampling procedure for selecting the largest multinomial cell probability," *Communications in Statistics-Simulation and Computation*, vol. 17, no. 3, pp. 969–994, 1988.
- [41] P. Chen and L. Hsu, "A composite stopping rule for multinomial subset selection," *British Journal of Mathematical and Statistical Psychology*, vol. 44, no. 2, pp. 403–411, 1991.
- [42] J. Miller, B. L. Nelson, and C. H. Reilly, "Efficient multinomial selection in simulation," *Naval Research Logistics (NRL)*, vol. 45, no. 5, pp. 459–482, 1998.
- [43] H. Vieira Jr, K. H. Kienitz, and M. C. N. Belderrain, "Multinomial selection problem: A study of bem and avc algorithms," *Communications in Statistics—Simulation and Computation®*, vol. 39, no. 5, pp. 971–980, 2010.
- [44] S. Chih Tsai and J. Wei, "Multinomial selection for comparison with a standard," *Communications in Statistics-Simulation and Computation*, vol. 40, no. 6, pp. 926–947, 2011.
- [45] R. G. Bechhofer, *Design and analysis of experiment for statistical selection, screening, and multiple comparisons*, 04; QA279, B4. 1995.
- [46] E. Tollefson, D. Goldsman, A. J. Kleywegt, and C. A. Tovey, "A comparative study of procedures for the multinomial selection problem," in *Essays in Production, Project Planning and Scheduling*, Springer, 2014, pp. 123–160.

- [47] H. Vieira Jr, S. M. Sanchez, P. J. Sanchez, K. H. Kienitz, and M. C. N. Belderrain, “A restricted multinomial hybrid selection procedure,” *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 24, no. 2, pp. 1–22, 2014.
- [48] H. Kesten and N. Morse, *A property of the multinomial distribution*. United States Air Force, Office of Scientific Research, 1958.
- [49] G. Chowell and H. Nishiura, “Characterizing the transmission dynamics and control of ebola virus disease,” *PLoS biology*, vol. 13, no. 1, e1002057, 2015.
- [50] N. C. Grassly and C. Fraser, “Mathematical models of infectious disease transmission,” *Nature Reviews Microbiology*, vol. 6, no. 6, p. 477, 2008.
- [51] C. Metcalf, W. Edmunds, and J Lessler, “Six challenges in modelling for public health policy,” *Epidemics*, vol. 10, pp. 93–96, 2015.
- [52] O. Diekmann and J. A. P. Heesterbeek, *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. John Wiley & Sons, 2000, vol. 5.
- [53] P. E. Fine and I. A. Carneiro, “Transmissibility and persistence of oral polio vaccine viruses: Implications for the global poliomyelitis eradication initiative,” *American journal of epidemiology*, vol. 150, no. 10, pp. 1001–1021, 1999.
- [54] W. E. R. Team, “Ebola virus disease in west africa—the first 9 months of the epidemic and forward projections,” *New England Journal of Medicine*, vol. 371, no. 16, pp. 1481–1495, 2014.
- [55] J. Li, D. Blakeley, *et al.*, “The failure of r_0 ,” *Computational and mathematical methods in medicine*, vol. 2011, 2011.
- [56] E. M. Volz, J. C. Miller, A. Galvani, and L. A. Meyers, “Effects of heterogeneous and clustered contact patterns on infectious disease dynamics,” *PLoS computational biology*, vol. 7, no. 6, e1002042, 2011.
- [57] Z. Mukandavire, S. Liao, J. Wang, H. Gaff, D. L. Smith, and J. G. Morris, “Estimating the reproductive numbers for the 2008–2009 cholera outbreaks in zimbabwe,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 21, pp. 8767–8772, 2011.
- [58] C. Caminade, J. Turner, S. Metelmann, J. C. Hesson, M. S. Blagrove, T. Solomon, A. P. Morse, and M. Baylis, “Global risk model for vector-borne transmission of zika virus reveals the role of el niño 2015,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 1, pp. 119–124, 2017.

- [59] T. A. Perkins, A. S. Siraj, C. W. Ruktanonchai, M. U. Kraemer, and A. J. Tatem, “Model-based projections of zika virus infections in childbearing women in the americas,” *Nature microbiology*, vol. 1, no. 9, p. 16 126, 2016.
- [60] H. Heesterbeek, R. M. Anderson, V. Andreasen, S. Bansal, D. De Angelis, C. Dye, K. T. Eames, W. J. Edmunds, S. D. Frost, S. Funk, *et al.*, “Modeling infectious disease dynamics in the complex landscape of global health,” *Science*, vol. 347, no. 6227, aaa4339, 2015.
- [61] F. M. Guerra, S. Bolotin, G. Lim, J. Heffernan, S. L. Deeks, Y. Li, and N. S. Crowcroft, “The basic reproduction number (r_0) of measles: A systematic review,” *The Lancet Infectious Diseases*, vol. 17, no. 12, e420–e428, 2017.
- [62] S. Bansal, B. T. Grenfell, and L. A. Meyers, “When individual behaviour matters: Homogeneous and network models in epidemiology,” *Journal of the Royal Society Interface*, vol. 4, no. 16, pp. 879–891, 2007.
- [63] S Davis, P Trapman, H. Leirs, M. Begon, and J. Heesterbeek, “The abundance threshold for plague as a critical percolation phenomenon,” *Nature*, vol. 454, no. 7204, p. 634, 2008.
- [64] M. J. Keeling, “The effects of local spatial structure on epidemiological invasions,” *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 266, no. 1421, pp. 859–867, 1999.
- [65] M. Salathé and J. H. Jones, “Dynamics and control of diseases in networks with community structure,” *PLoS computational biology*, vol. 6, no. 4, e1000736, 2010.
- [66] Q.-H. Liu, M. Ajelli, A. Aleta, S. Merler, Y. Moreno, and A. Vespignani, “Measurability of the epidemic reproduction number in data-driven contact networks,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 50, pp. 12 680–12 685, 2018.
- [67] E Goldstein, K Paur, C Fraser, E Kenah, J Wallinga, and M Lipsitch, “Reproductive numbers, epidemic spread and control in a community of households,” *Mathematical biosciences*, vol. 221, no. 1, pp. 11–25, 2009.
- [68] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz, “Superspreading and the effect of individual variation on disease emergence,” *Nature*, vol. 438, no. 7066, p. 355, 2005.
- [69] M. Laumanns and R. Zenklusen, “Computational complexity of impact size estimation for spreading processes on networks,” *The European Physical Journal B*, vol. 71, no. 4, p. 481, 2009.

- [70] M. Shapiro and E. Delgado-Eckert, “Finding the probability of infection in an sir network is np-hard,” *Mathematical biosciences*, vol. 240, no. 2, pp. 77–84, 2012.
- [71] V. V. Vazirani, *Approximation algorithms*. Springer Science & Business Media, 2013.
- [72] LandRussia, *Population density by municipal divisions*, https://www.mapmania.org/map/69809/density_of_population_in_russia_in_2013, 2018.
- [73] WHO, *The Ministry of Public Health and Population announces cholera cases in Yemen*, <http://www.emro.who.int/media/news/the-ministry-of-health-announces-cholera-cases-in-yemen.html>, 2016.
- [74] F.-X. Weill, D. Domman, E. Njamkepo, A. A. Almesbahi, M. Naji, S. S. Nasher, A. Rakesh, A. M. Assiri, N. C. Sharma, S. Kariuki, *et al.*, “Genomic insights into the 2016–2017 cholera epidemic in yemen,” *Nature*, p. 1, 2019.
- [75] WHO, *Yemen - Cholera Outbreak Epidemiology Update Data*, <https://data.humdata.org/dataset/yemen-cholera-outbreak-daily-epidemiology-update>, 2018.
- [76] M. C. Eisenberg, S. L. Robertson, and J. H. Tien, “Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease,” *Journal of theoretical biology*, vol. 324, pp. 84–102, 2013.
- [77] D. M. Hartley, J. G. Morris Jr, and D. L. Smith, “Hyperinfectivity: A critical element in the ability of v. cholerae to cause epidemics?” *PLoS medicine*, vol. 3, no. 1, e7, 2005.
- [78] Z. Mukandavire, D. L. Smith, and J. G. Morris Jr, “Cholera in haiti: Reproductive numbers and vaccination coverage estimates,” *Scientific reports*, vol. 3, p. 997, 2013.
- [79] M. Eppinger, T. Pearson, S. S. Koenig, O. Pearson, N. Hicks, S. Agrawal, F. Sanjar, K. Galens, S. Daugherty, J. Crabtree, *et al.*, “Genomic epidemiology of the haitian cholera outbreak: A single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic,” *MBio*, vol. 5, no. 6, e01721–14, 2014.
- [80] WorldPop, University of Southampton, UK, *Yemen - population*, <http://www.worldpop.org.uk/data/summary/?doi=10.5258/SOTON/WP00303>, 2013.

- [81] Yemen Central Statistical Organisation, *Yemen - Water Courses*, <https://data.humdata.org/dataset/yemen-water-bodies>, 2014.
- [82] M. L. Bennish, “Cholera: Pathophysiology, clinical features, and treatment,” in *Vibrio cholerae and Cholera*, American Society of Microbiology, 1994, pp. 229–255.
- [83] V. Tudor, I. Strati, *et al.*, *Smallpox: cholera*. 1977.
- [84] Yemen Central Statistical Organization, *Yemen: CSO 2017 Population Projections by Governorate and District, Sex and Age Disaggregated*, <https://data.humdata.org/dataset/yemen-cso-2017-population-projections-by-governorate-district-sex-age-disaggregated>, 2017.
- [85] ———, *Yemen - Administrative Boundaries*, <https://data.humdata.org/dataset/yemen-admin-boundaries>, 2017.
- [86] International Organization for Migration (IOM), Displacement Tracking Matrix (DTM), *Yemen displacement data - task force on population movement*, <https://data.humdata.org/dataset/yemen-displacement-data-task-force-on-population-movement-iom-dtm>, 2018.
- [87] M. Pascual, X. Rodó, S. P. Ellner, R. Colwell, and M. J. Bouma, “Cholera dynamics and el niño-southern oscillation,” *Science*, vol. 289, no. 5485, pp. 1766–1769, 2000. eprint: <http://science.sciencemag.org/content/289/5485/1766.full.pdf>.
- [88] Armed Conflict Location and Event Data Project (ACLED), *Conflict data for yemen*, <https://data.humdata.org/dataset/acled-data-for-yemen>, 2019.
- [89] R. Cohen, S. Havlin, and D. Ben-Avraham, “Efficient immunization strategies for computer networks and populations,” *Physical review letters*, vol. 91, no. 24, p. 247 901, 2003.
- [90] J. L. Ritchie-Dunham and J. F. Méndez Galván, “Evaluating epidemic intervention policies with systems thinking: A case study of dengue fever in mexico,” *System Dynamics Review: The Journal of the System Dynamics Society*, vol. 15, no. 2, pp. 119–138, 1999.
- [91] M. M. Engelgau, C. H. Woernle, R. T. Rolfs, J. R. Greenspan, M. O’Cain, and R. D. Gorsky, “Control of epidemic early syphilis: The results of an intervention campaign using social networks.,” *Sexually transmitted diseases*, vol. 22, no. 4, pp. 203–209, 1995.

- [92] Z. Wang, C. T. Bauch, S. Bhattacharyya, A. d’Onofrio, P. Manfredi, M. Perc, N. Perra, M. Salathé, and D. Zhao, “Statistical physics of vaccination,” *Physics Reports*, vol. 664, pp. 1–113, 2016.
- [93] M. J. Ferrari, S. Bansal, L. A. Meyers, and O. N. Bjørnstad, “Network frailty and the geometry of herd immunity,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 273, no. 1602, pp. 2743–2748, 2006.
- [94] J. Hindes and I. B. Schwartz, “Epidemic extinction and control in heterogeneous networks,” *Physical review letters*, vol. 117, no. 2, p. 028 302, 2016.
- [95] J. Marcelino and M. Kaiser, “Reducing influenza spreading over the airline network.,” *PLoS currents*, vol. 1, RRN1005–RRN1005, 2009.
- [96] J. Omić, J. Martín-Hernández, and P. Van Mieghem, “Network protection against worms and cascading failures using modularity partitioning,” in *2010 22nd International Teletraffic Congress (ITC 22)*, IEEE, 2010, pp. 1–8.
- [97] T. Easton, K. Carlyle, and C. Scoglio, “Optimizing quarantine regions through ellipsoidal geographic networks,” *Computers & Industrial Engineering*, vol. 80, pp. 145–153, 2015.
- [98] M. L. Cohen, “Changing patterns of infectious disease,” *Nature*, vol. 406, no. 6797, pp. 762–767, 2000.
- [99] J. C. Miller and J. M. Hyman, “Effective vaccination strategies for realistic social networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 386, no. 2, pp. 780–785, 2007.
- [100] B. Vidondo, M. Schwehm, A. Bühlmann, and M. Eichner, “Finding and removing highly connected individuals using suboptimal vaccines,” *BMC infectious diseases*, vol. 12, no. 1, p. 51, 2012.
- [101] S. P. Borgatti, “Centrality and network flow,” *Social networks*, vol. 27, no. 1, pp. 55–71, 2005.
- [102] J. G. Restrepo, E. Ott, and B. R. Hunt, “Characterizing the dynamical importance of network nodes and links,” *Physical review letters*, vol. 97, no. 9, p. 094 102, 2006.
- [103] C. M. Schneider, T. Mihaljev, S. Havlin, and H. J. Herrmann, “Suppressing epidemics with a limited amount of immunization units,” *Physical Review E*, vol. 84, no. 6, p. 061 911, 2011.

- [104] C. M. Schneider, T. Mihaljev, and H. J. Herrmann, “Inverse targeting—an effective immunization strategy,” *EPL (Europhysics Letters)*, vol. 98, no. 4, p. 46 002, 2012.
- [105] M. Zarghamy and V. Preciadoy, “Worst-case scenarios for greedy, centrality-based network protection strategies,” in *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, IEEE, 2014, pp. 1–6.
- [106] X.-L. Ren, N. Gleinig, D. Helbing, and N. Antulov-Fantulin, “Generalized network dismantling,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 14, pp. 6554–6559, 2019.
- [107] J. Aspnes, K. Chang, and A. Yampolskiy, “Inoculation strategies for victims of viruses and the sum-of-squares partition problem,” *Journal of Computer and System Sciences*, vol. 72, no. 6, pp. 1077–1093, 2006.
- [108] A. Braunstein, L. Dall’Asta, G. Semerjian, and L. Zdeborová, “Network dismantling,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 44, pp. 12 368–12 373, 2016.
- [109] E. A. Enns, J. J. Mounzer, and M. L. Brandeau, “Optimal link removal for epidemic mitigation: A two-way partitioning approach,” *Mathematical biosciences*, vol. 235, no. 2, pp. 138–147, 2012.
- [110] T. Gross, C. J. D. D’Lima, and B. Blasius, “Epidemic dynamics on an adaptive network,” *Physical review letters*, vol. 96, no. 20, p. 208 701, 2006.
- [111] S. Finbow and G. MacGillivray, “The firefighter problem: A survey of results, directions and questions.,” *Australasian J. Combinatorics*, vol. 43, pp. 57–78, 2009.
- [112] D. S. Altner, Ö. Ergun, and N. A. Uhan, “The maximum flow network interdiction problem: Valid inequalities, integrality gaps, and approximability,” *Operations Research Letters*, vol. 38, no. 1, pp. 33–38, 2010.
- [113] K. J. Cormican, D. P. Morton, and R. K. Wood, “Stochastic network interdiction,” *Operations Research*, vol. 46, no. 2, pp. 184–197, 1998.
- [114] P. Vanhems, A. Barrat, C. Cattuto, J.-F. Pinton, N. Khanafer, C. Régis, B.-a. Kim, B. Comte, and N. Voirin, “Estimating potential infection transmission routes in hospital wards using wearable proximity sensors,” *PloS one*, vol. 8, no. 9, 2013.
- [115] L. Ozella, F. Gesualdo, M. Tizzoni, C. Rizzo, E. Pandolfi, I. Campagna, A. E. Tozzi, and C. Cattuto, “Close encounters between infants and household members measured through wearable proximity sensors,” *PloS one*, vol. 13, no. 6, 2018.

- [116] A. E. Aiello, A. M. Simanek, M. C. Eisenberg, A. R. Walsh, B. Davis, E. Volz, C. Cheng, J. J. Rainey, A. Uzicanin, H. Gao, *et al.*, “Design and methods of a social network isolation study for reducing respiratory infection transmission: The ex-flu cluster randomized trial,” *Epidemics*, vol. 15, pp. 38–55, 2016.
- [117] A. J. Kleywegt, A. Shapiro, and T. Homem-de Mello, “The sample average approximation method for stochastic discrete optimization,” *SIAM Journal on Optimization*, vol. 12, no. 2, pp. 479–502, 2002.
- [118] R. K. Wood, “Deterministic network interdiction,” *Mathematical and Computer Modelling*, vol. 17, no. 2, pp. 1–18, 1993.
- [119] A. A. Ageev and M. I. Sviridenko, “Pipage rounding: A new method of constructing algorithms with proven performance guarantee,” *Journal of Combinatorial Optimization*, vol. 8, no. 3, pp. 307–328, 2004.
- [120] L. Gurobi Optimization, “Gurobi optimizer 8.0,” *Beaverton, OR: Gurobi Optimization LLC*, 2018.
- [121] G. Laporte and F. V. Louveaux, “The integer l-shaped method for stochastic integer programs with complete recourse,” *Operations research letters*, vol. 13, no. 3, pp. 133–142, 1993.
- [122] W. W. Zachary, “An information flow model for conflict and fission in small groups,” *Journal of anthropological research*, vol. 33, no. 4, pp. 452–473, 1977.
- [123] M. C. Kiti, M. Tizzoni, T. M. Kinyanjui, D. C. Koech, P. K. Munywoki, M. Meriac, L. Cappa, A. Panisson, A. Barrat, C. Cattuto, *et al.*, “Quantifying social contacts in a household setting of rural kenya using wearable proximity sensors,” *EPJ data science*, vol. 5, no. 1, pp. 1–21, 2016.
- [124] J. Fournet and A. Barrat, “Contact patterns among high school students,” *PloS one*, vol. 9, no. 9, 2014.
- [125] M. Girvan and M. E. Newman, “Community structure in social and biological networks,” *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.
- [126] W. Hoeffding, “Probability inequalities for sums of bounded random variables,” in *The Collected Works of Wassily Hoeffding*, Springer, 1994, pp. 409–426.
- [127] B. Bollobás, *Modern graph theory*. Springer Science & Business Media, 2013, vol. 184.

- [128] I. Benjamini, S. Boucheron, G. Lugosi, R. Rossignol, *et al.*, “Sharp threshold for percolation on expanders,” *The Annals of Probability*, vol. 40, no. 1, pp. 130–145, 2012.
- [129] R. M. Karp, “The transitive closure of a random digraph,” *Random Structures & Algorithms*, vol. 1, no. 1, pp. 73–93, 1990.