University of Pennsylvania
ScholarlyCommons

Publicly Accessible Penn Dissertations

2019

Testing Infection Graphs

Justin Turner Khim
University of Pennsylvania, jtkhim@gmail.com

Follow this and additional works at: https://repository.upenn.edu/edissertations

 Part of the Statistics and Probability Commons

Recommended Citation
Khim, Justin Turner, "Testing Infection Graphs" (2019). *Publicly Accessible Penn Dissertations*. 3358.
https://repository.upenn.edu/edissertations/3358

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/3358
For more information, please contact repository@pobox.upenn.edu.

Testing Infection Graphs

Abstract
We study the following problem: given two graphs G_0 and G_1 defined on a common set of n vertices and a single observation of the statuses of these vertices, i.e. either infected, uninfected, or censored, did the infection spread on G_0 or G_1? Modern instances of such ``infections'' include diseases such as HIV, behaviors such as smoking, or information such as online news articles. For particular stochastic spreading mechanisms, we give algorithms for this testing problem based on hypothesis discretization and permutation-invariance. Additionally, these methods also lead to confidence sets for parameters that also govern the spread of infection and for the graphs on which the infection spread.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Statistics

First Advisor
Zongming Ma

Keywords
automorphisms, discretization, hypothesis testing, social and information networks

Subject Categories
Statistics and Probability

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/3358

TESTING INFECTION GRAPHS

Justin Turner Khim

A DISSERTATION
in

Statistics

For the Graduate Group in Managerial Science and Applied Economics
Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
2019

Supervisor of Dissertation

Zongming Ma
Associate Professor of Statistics

Graduate Group Chairperson

Catherine Schrand
Celia Z. Moh Professor, Professor of Accounting

Dissertation Committee:

Dylan Small
Class of 1965 Wharton Professor of Statistics

Alexander Rakhlin
Associate Professor of Brain & Cognitive Sciences
Massachusetts Institute of Technology

Po-Ling Loh
Assistant Professor of Statistics
University of Wisconsin-Madison

TESTING INFECTION GRAPHS
© COPYRIGHT 2019
Justin Turner Khim

This work is licensed under the
Creative Commons Attribution
NonCommercial-ShareAlike 3.0
License
To view a copy of this license, visit
http://creativecommons.org/licenses/by-nc-sa/3.0/

Acknowledgments:

I would like to thank my family, friends, and teachers for being more or less supportive in getting me to this point.

I would also like to thank my advisor, Po-Ling Loh, for somehow managing to put up with me.

ABSTRACT

TESTING INFECTION GRAPHS

Justin Khim

Zongming Ma

We study the following problem: given two graphs $\mathcal{G}_0$ and $\mathcal{G}_1$ defined on a common set of n vertices and a single observation of the statuses of these vertices, i.e. either infected, uninfected, or censored, did the infection spread on $\mathcal{G}_0$ or $\mathcal{G}_1$? Modern instances of such "infections" include diseases such as HIV, behaviors such as smoking, or information such as online news articles. For particular stochastic spreading mechanisms, we give algorithms for this testing problem based on hypothesis discretization and permutation-invariance. Additionally, these methods also lead to confidence sets for parameters that also govern the spread of infection and for the graphs on which the infection spread.

Contents

List of Tables

List of Figures

1. Introduction

Information, diseases, and the adoption of certain behaviors may spread according to a network of relationships connecting susceptible individuals (Anderson and May, 1992; Christakis and Fowler, 2007; Jackson, 2008; Morris, 1993; Newman, 2002). Natural questions to address include predicting the pathway or scope of a disease; inferring the source; and identifying nodes or edges at which to intervene in order to slow the spread of the epidemic. The answers to these questions may vary depending on the stochastic mechanism governing the spread of disease between individuals.

Algorithms for addressing such questions often assume knowledge of the underlying graph structure (Borgs et al., 2012; Brautbar and Kearns, 2010; Bubeck et al., 2017). For instance, in the influence maximization problem, the goal is to identify an optimal set of nodes to initially infect in order to propagate a certain behavior as widely as possible (Domingos and Richardson, 2001; Kempe et al., 2003). Due to submodularity of the influence function, one may obtain a constant factor approximation to the influence-maximizing subset using a simple greedy algorithm. However, both the connectivity of the network and edge transmission parameters must be known in order to successfully execute the algorithm (Chen et al., 2010). Similarly, algorithms for network immunization, which aim to eliminate an epidemic by performing targeted interventions, assume knowledge of the graph (Albert et al., 2000; Cohen et al., 2003; Pastor-Satorras and Vespignani, 2002). In real-world applications, however, prior knowledge of the edge structure of the underlying network or parameters of the infection spreading mechanism may be unavailable.

Accordingly, the focus of our paper is the inference problem of identifying the underlying network based on observed infection data. One approach involves employing tools from graphical model estimation, since the graphical model corresponding to joint vectors of infection times coincides with the unknown network (Netrapalli and Sanghavi, 2012; Gomez-Rodriguez et al., 2016). However, these methods critically leverage the availability of time-stamped data and observations of multiple i.i.d. infection processes spreading over the same graph. Another line of work concerns reconstructing an infection graph based on the order in which nodes are infected over multiple infection processes, and provides bounds on the number of distinct observations required to recover the edge structure of the graph. These methods are attractive in that they do not assume a particular stochastic spreading model, and are even guaranteed to reconstruct the true network when the observations are chosen adversarially (Angluin et al., 2015; Huang et al., 2017). On the other hand, data from multiple infections are still assumed to be available.

In contrast to these papers, we are interested in studying scenarios where infection data are available for a *single* snapshot of a single epidemic outbreak. For instance, if the goal is to perform optimal interventions on a network of individuals during an outbreak such as Ebola (Dudas et al., 2017), data may only be available about the infection status of individuals at a given point in time. Although historic data may have been collected concerning the spread of other epidemics on the same network, there is no guarantee that the other diseases will

spread according to the same mechanisms, and the network may have changed over time. Instead of employing the aforementioned graph estimation procedures based on observation vectors, we cast the problem in the form of a hypothesis test: Given two candidate graphs, our goal is to identify the graph on which the infection has propagated. This type of problem was previously studied by Milling et al. (2015), who proposed inference procedures for testing an empty graph versus graphs satisfying "speed and spread" conditions, and for testing two graphs against each other, critically using an "independent neighborhoods" assumption. A graph testing problem of a somewhat similar flavor may also be found in Bubeck et al. (2016), but their goal is to identify the model of network formation for a random graph, rather than determine the network underlying an epidemic outbreak.

In this dissertation, we primarily examine two approaches: discretization and permutation. Discretization of the null hypothesis space can be thought of as a computational solution to a particular non-monotonicity for composite hypotheses. Perhaps more elegantly, our permutation approach leverages specific symmetries, vastly reducing computation and leading to permutation tests. We study this approach in more detail.

Permutation testing is notion which dates back to Fisher (1935). Permutation tests are classically applied when random variables are exchangeable under the null hypothesis. The test statistic is then computed with respect to random permutations of the observation vector, and the observed statistic is compared to quantiles of the simulated distribution (Good, 2013). We adapt this technique in a novel manner to the graph testing problem. The procedure is the same as in classical permutation testing, where we recompute a certain statistic on a set of randomly chosen permutations of the infection "vector" recording the observed statuses of the nodes in the graph. Based on quantiles of the empirical histogram, we calculate a rejection rule for the test.

The key idea is that if the null hypothesis corresponds to an empty or complete graph and edge transmission rates are homogeneous, the components of the infection vector are certainly exchangeable. This is an important setting in its own right, since it allows us to determine whether a network structure describes an infection better than random noise. However, the permutation test succeeds more generally under appropriate assumptions incorporating symmetries of both the null graph and alternative graph.

A scientific motivation for graph hypothesis testing is a case when a practitioner wants to decide whether a disease is spreading according to a fixed hypothesized graph structure versus completely random transmissions, which would correspond to an empty graph. Another plausible scenario is when a scientist wants to test a long-standing hypothesis that a genetic network follows a particular connectivity pattern, versus a proposed alternative involving the addition or removal of certain edges. In a third setting, one might encounter two distinct graphs representing connections between individuals on different social network platforms, and hypothesize that information diffusion is governed by one graph instead of the other.

This dissertation is organized into three main parts: the preliminary material, results on discretization, and results on permutations. The latter two parts are more substantial and have separate chapters for the primary results, proofs and secondary theoretical results, and numerical results.

Part I

Preliminaries

2. Setup

In this chapter, we introduce our stochastic models and our testing problem. First though, we provide some preliminaries on graph theory and a visual example that we shall return to throughout our numerical analyses.

2.1 Basic Graph Terminology

The primary mathematical object of interest for us is a graph. A *graph* $\mathcal{G}$ is a pair of sets $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the elements of $\mathcal{V}$ are called *vertices* and the elements of $\mathcal{E}$ are called *edges*. For simplicity, we consider the vertices $\mathcal{V} = \{1, \ldots, n\}$. Now, an edge e is a pair of vertices $e = (u, v)$. If the edges are unordered, then the resulting graph is called an undirected graph. Otherwise, the graph is a directed graph. For our results, we shall primarily deal with undirected graphs, although for our discretization results, the graphs may be directed without changing the results. Note that we use the words "network" and "graph" interchangeably.

Additionally, we are interested in considering edge weights for our graphs. Given a function $w : \mathcal{E} \to \mathbb{R}_+$, we say that $w(e)$ is the *weight* of an edge e. If a graph $\mathcal{G}$ has an edge weight function w, then we say that $\mathcal{G}$ is a graph with weighted edges. For our permutation results, we shall restrict ourselves to unweighted graphs, which we can think of as $w(e) = 1$ for an edge e.

Finally, one last standard definition is the cut of a graph. A *cut* C is a partition of the vertices into two sets $C = (U, V)$. The cut-set consists of the edges with one vertex in U and one vertex in V. A maximum cut, usually abbreviated maxcut, is a cut that maximizes the weight of the cut.

2.2 A Visual Example

An ongoing example that we consider when applying our analyses is an HIV network. The network is constructed from the population of Colorado Springs from 1982–1989, and the observed infection statuses of individuals, obtained from the CDC and the HIV Counseling and Testing Center, is reported in Potterat et al. (2002). Here, the infection statuses are infected, uninfected, and censored, meaning that the HIV infection status was not observed. For a visualization of the network, see Figure 2.1. The network was constructed from contact tracing for sexual and injecting drug partners; beginning in 1985, officials interviewed people testing positive for HIV for information to identify relevant partners.

In the case of this HIV network, we can think of graph testing as verifying the importance of the proposed network structure and evaluating possible competing network structures. From the network construction, it is possible that there were in fact additional edges between the censored vertices, and if so perhaps the proposed network structure would not be sufficient

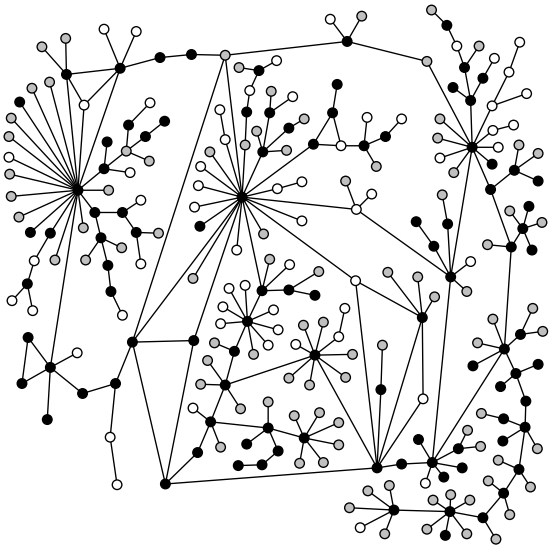

Figure 2.1 The HIV infection graph. Black vertices are infected, white vertices are uninfected, and gray vertices are censored. Edges represent sexual or injecting drug use relationships. Adapted by permission from BMJ Publishing Group, Ltd. [*Sexually Transmitted Infections*, Potterat et al., 78, i159–i163, 2002]

for explaining the infection. It is quite easy to imagine this in other settings, such as online social networks. In particular, Myers et al. (2012) observed that in such settings, many infections do not occur over the edges of the observed network.

2.3 Infection Notation

Let $\mathcal{G}_0 = (\mathcal{V}, \mathcal{E}_0)$ and $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1)$ denote two graphs defined over a common set of vertices $\mathcal{V} = \{1, \ldots, n\}$. We define weight functions $w_0 : \mathcal{E}_0 \to \mathbb{R}_+$ and $w_1 : \mathcal{E}_1 \to \mathbb{R}_+$, where $w_i(e)$ is the weight of the edge e in $\mathcal{G}_i$.

A random vector of infection statuses $\mathcal{I} := \{\mathcal{I}_v\}_{v \in \mathcal{V}}$ consists of entries

$$\mathcal{I}_v = \begin{cases} 1, & \text{if } v \text{ is infected,} \\ 0, & \text{if } v \text{ is uninfected,} \\ \star, & \text{if } v \text{ is censored.} \end{cases}$$

Let $\mathcal{I}^1$ denote the set of infected vertices, and let $\mathcal{I}^0$ and $\mathcal{I}^\star$ be defined analogously. We then define the space of possible infection status vectors involving exactly k infected vertices, c censored vertices, and $n - k - c$ uninfected vertices:

$$\mathbb{I}_{k,c} = \Big\{ J \in \{0, 1, \star\}^n : |J^1| = k,\ |J^\star| = c \Big\}.$$

2.4 Models

In this section, we introduce our main spreading model. We also define an Ising model, against which we compare our spreading model in a few situations. Finally, we discuss a general methods of censoring that may be applied in either model.

2.4.1 Stochastic Spreading Model

Our main infection model includes the following components: spreading by contagion along edges of the network, and spreading via factors external to the network. The rates of spreading are governed by the nonnegative parameter $\beta \in \mathbb{R}_+$. We assume that an infection spreads on $\mathcal{G}$, beginning at time 0, as follows:

(i) For each vertex v, generate an independent random variable $T_v \sim \text{Exp}(1)$.

(ii) For each edge $(u, v) \in \mathcal{E}$, generate an independent random variable $T_{uv} \sim \text{Exp}(\beta w(u, v))$ where w is the weight function.

(iii) For each vertex v, define the infection time $t_v := \min_{u \in N(v)}\{t_u + T_{uv}\} \wedge T_v$, where $N(v)$ is the set of neighbors of v.

In other words, each vertex v contracts the disease via random infection according to an exponential random variable with rate 1, and contracts the disease from an infected neighbor u at rate $\beta w(u, v)$ relative to random infection; in particular, when $\mathcal{G}$ is an empty graph, each successively infected node is chosen uniformly at random. The presence of the random variable T_v has the interpretation that some factors involved in spreading the disease may not be fully accounted for in the hypothesized graphs. We define t_k to be the time at which the k^{th} reporting node becomes infected, and we suppose we observe the state of the graph at some time t such that $t_k \leq t \leq t_{k+1}$. Note that this differs from other settings, where the time t at which the node infection statuses are observed is assumed to be a known constant (Kesten, 1993; Milling et al., 2015; Shah and Zaman, 2011).

In our analysis, we rely on the notion of *paths*. A path $P = (P_1, \ldots, P_k)$ is an ordered set of vertices, and vertex P_i is the i^{th} vertex to be infected. A given infection J in $\mathbb{I}_{k,0}$ corresponds to $k!$ possible paths, which we collect into a set $\mathfrak{P}(J)$. We also refer to the path random variable as $\mathcal{P}$. Thus, we have the relation

$$\mathbb{P}(\mathcal{I} = J) = \sum_{P \in \mathfrak{P}(J)} \mathbb{P}(\mathcal{P} = P),$$

which we will use in our derivations. Additionally, we denote $P_{1:t} = (P_1, \ldots, P_t)$.

To understand the notation we shall use in computing path probabilities, suppose k vertices have been infected in running the process at time t, and we want to determine the $(k+1)^{\text{th}}$ infected vertex. Since all of the T_v's and T_{uv}'s are independent exponential random variables, the probability that vertex v is infected next is

$$\mathbb{P}(v \text{ is infected next} | \mathcal{P}_{1:t-1} = P_{1:t-1}) = \frac{1 + \beta W_{t,v}(P_{1:t-1})}{(n-k) + \beta W_t(P_{1:t-1})},$$

where $W_t(P_{1:t-1})$ is the sum of the weights of edges with one infected vertex at time t, and $W_{t,v}(P_{1:t-1})$ is the sum of the weights $w(u,v)$ where u is infected.

We make a few additional remarks on notation. First, when the edges are unweighted, we use N_t and $N_{t,v}$ instead of W_t and $W_{t,v}$ to emphasize the the count of neighbors is not weighted. Finally, if P is a path of length t' and $t < t'$, $N_{t,\text{in}}(P)$ is equivalent to $N_{t,P_t}(P_{1:t-1})$.

Now, note that we consider the case $\beta = \infty$. Here, uniform spread only occurs when there are no edges between infected and uninfected vertices, i.e. when $W_t = 0$. One theoretical benefit is that this allows for the compactness of our parameter space and space of distributions, although this is not necessary for our results.

Finally, we denote an infection $\mathcal{I}$ over a graph $\mathcal{G}$ with parameter β by $\mathcal{I} \sim \text{SSM}(\mathcal{G}, \beta)$. This simplifies the statements of our hypothesis tests.

2.4.2 Ising Model

We now describe an Ising model parametrized by $\beta \in \mathbb{R}_+$. Define the energy function $\mathcal{H} : \mathbb{I}_{k,0} \to \mathbb{R}$ by

$$\mathcal{H}(J) := -\sum_{(u,v)\in\mathcal{E}} w(u,v)J_uJ_v = -W(J),$$

where W is the edges-within statistic that we discuss in further detail in Section 2.6. We define the uncensored Ising model by

$$\mathbb{P}^u_{\text{Ising}}(\mathcal{I} = J) := \frac{\exp(-\beta\mathcal{H}(J))}{\sum_{J'\in\mathbb{I}_{k,0}} \exp(-\beta\mathcal{H}(J'))} = \frac{1}{Z_{\text{Ising},k}(\beta)} \exp(-\beta\mathcal{H}(J)).$$

Note that this resembles an Ising model with -1 replaced by 0. For simplicity, we denote an infection $\mathcal{I}$ generated by the Ising model on graphset $\mathcal{G}$ with parameter β by $\mathcal{I} \sim \text{Ising}(\mathcal{G}, \beta)$.

Since the Ising model does not have a spreading interpretation, this is not our model of interest. However, in certain regimes, the stochastic spreading model behaves similarly to the Ising model. Finally, at other points, the Ising model provides a stark contrast to our stochastic spreading model.

2.4.3 Censoring

Our model permits vertex statuses to be censored. The two models of censoring we consider are uniform censoring and fixed censoring. We consider the uniform spreading model here as an example, and we further study the fixed censoring variant in Chapter 8, since the precise details are most important for our permutation results.

Our goal is to condition on c censored vertices, which are chosen uniformly. Given an infection vector J that may contain censored components, we define the set $\mathcal{U}(J)$ to consist of all infection vectors J' such that J' has no censored vertices and $J_v = J'_v$ for each

uncensored vertex v. We then define the measure

$$\mu(J;\beta) = \sum_{J' \in \mathcal{U}(J)} \frac{1}{\binom{n}{c}} \mathbb{P}^u(\mathcal{I} = J'),$$

where $\mathbb{P}^u(\mathcal{I} = J')$ denotes the probability of an uncensored infection from the model described above. In other words, μ simply computes the probability of observing a vector J after randomly censoring c nodes. If we define the normalizing constant

$$Z(\beta) := \sum_{J \in \mathbb{I}_{k,c}} \mu(J;\beta) = \sum_{J \in \mathbb{I}_{k,c}} \sum_{J' \in \mathcal{U}(J)} \frac{1}{\binom{n}{c}} \mathbb{P}^u(\mathcal{I} = J'),$$

which provides the probability of observing any vector with c censored nodes, we see that

$$\mathbb{P}(\mathcal{I} = J) = \frac{1}{Z(\beta)} \mu(J;\beta)$$

is the probability of observing infection vector J, conditioned on c nodes having been censored. Note that in the case when $\mathcal{G}$ is the empty graph, the values of $\mu(J;\beta)$ are the same for all values of J, so $\mathbb{P}(\mathcal{I}) = 1/|\mathbb{I}_{k,c}| = 1/\binom{n}{k\ c}$, i.e. the inverse of the multinomial coefficient for choosing k and c items from a set of n.

2.5 Parameters and Tests

In this section, we describe our hypothesis testing and confidence set framework. To this end, we start by more fully detailing the notation of our parameters. Let $\mathcal{G}_0$ and $\mathcal{G}_1$ be graphs with possibly weighted edges. Let $\mathfrak{B}_0$ and $\mathfrak{B}_1$ be sets of possible β values, and these are subsets of the natural parameter spaces $\mathfrak{C}_0$ and $\mathfrak{C}_1$. A graph-parameter pair $(\mathcal{G}_i, \beta)$ parametrizes a probability measure $\mathbb{P}_{i,\beta}$ by $\mathcal{M}(\mathcal{G}_i, \beta) = \mathbb{P}_{i,\beta}$, and $\mathbb{P}_{i,\beta}$ can be thought of as an element of the probability simplex with entries indexed by the infections of $\mathbb{I}_{k,c}$. Note that we sometimes write our parameters as $\theta = (\mathcal{G}, \beta)$, and the resulting parameter space is then Θ.

Now, our hypothesis tests have the following form:

$$H_0 : \mathcal{I} \sim \mathrm{SSM}(\mathcal{G}_0, \beta_0) \text{ for } \beta_0 \in \mathfrak{B}_0,\ \mathcal{G}_0 \in \mathfrak{G}_0$$
$$H_1 : \mathcal{I} \sim \mathrm{SSM}(\mathcal{G}_1, \beta_1) \text{ for } \beta_1 \in \mathfrak{B}_1,\ \mathcal{G}_1 \in \mathfrak{G}_1.$$

In most sections, we only consider a "simple" hypothesis test where $\mathfrak{G}_0$ and $\mathfrak{G}_1$ are singleton sets. Via duality, we can define confidence sets as the set of $(\mathcal{G}_0, \beta_0)$ for which the test was not rejected.

At this point, we make some of these sets more concrete. Our primary hypothesis testing problem has

$$\mathfrak{C}_0 = \mathfrak{C}_1 = \overline{\mathbb{R}}_+ = [0, \infty].$$

In general, a reasonable choice of parameter sets might be $\mathfrak{B}_0 = [0, \infty]$ and $\mathfrak{B}_1 = (0, \infty]$. Note that if $0 \in \mathfrak{B}_0 \cap \mathfrak{B}_1$, the parameter spaces of the null and alternative hypotheses

overlap, so the minimum power of any valid test is at most α. Thus, to provably attain non-trivial power, one could take the standard approach of designating an indifference region, so that $\mathfrak{B}_1 = [c, \infty]$ for some $c > 0$.

A critical function $\psi : \mathbb{I}_{k,c} \to \{0, 1\}$ may be used to indicate the result of a test. We will be interested in bounding the risk, which is the sum of Type I and Type II errors:

$$R_{k,c}(\psi; \beta_0, \beta_1) = \mathbb{P}_{0,\beta_0}(\psi(\mathcal{I}) = 1) + \mathbb{P}_{1,\beta_1}(\psi(\mathcal{I}) = 0).$$

2.6 Test Statistics and Thresholds

In general, we would like to use any test statistic $S : \mathbb{I}_{k,c} \to \mathbb{R}$. However, in Part III when we consider permutations, we shall have to make some restrictions. Without loss of generality, we reject the null hypothesis for large values of S. An important statistic for our analyses is the edge weights within an infection

$$W(J) := \sum_{(u,v)\in\mathcal{E}} w(u, v)\mathbf{1}\{J_u = J_v = 1\}.$$

For a fixed value of β, we define the $(1 - \alpha)$-quantile of $S(\mathcal{I})$ to be

$$t_{\alpha,\beta} := \sup\{t \in \text{supp}(S) : \mathbb{P}_\beta(S(\mathcal{I}) \geq t) > \alpha\}$$

where $\text{supp}(S)$ is the support of S, i.e. the set of values that the discrete random variable S takes with positive probability. Note that we have the two bounds

$$\begin{aligned} \mathbb{P}_\beta(S(\mathcal{I}) > t_{\alpha,\beta}) &\leq \alpha \\ \mathbb{P}_\beta(S(\mathcal{I}) \geq t_{\alpha,\beta}) &> \alpha, \end{aligned} \tag{2.1}$$

i.e. the strictness of the inequality in the event determines the direction of the inequality in the probability bound.

The tests given in this dissertation are based on simulated null distributions of the test statistic S in order to approximate $t_{\alpha,\beta}$. To quantify the exact uncertainty introduced by the simulations, we define the empirical $(1 - \alpha)$-quantile $\hat{t}_\alpha : \mathbb{I}_{k,c}^{N_{\text{sims}}} \to \mathbb{R}$ to be

$$\hat{t}_{\alpha,\beta} := \sup\left\{t \in \text{supp}(S) : \frac{1}{N_{\text{sims}}} \sum_{i=1}^{N_{\text{sims}}} \mathbf{1}\{S(I_i) \geq t\} \geq \alpha\right\}.$$

Additional Notation

Before proceeding, we comment on a few of our notation conventions. As seen earlier in the case of paths, we use $1 : t$ to denote a t-tuple. Other examples include $\mathcal{I}^{(1:m)} = (\mathcal{I}^{(1)}, \ldots, \mathcal{I}^{(m)}$, where each $\mathcal{I}^{(i)}$ is an infection, and $X_{1:k+c} = (X_1, \ldots, X_{k+c})$.

For probability measures, we write $\mathbb{P}_{i,\beta}$ to denote a probability computed with respect to the parameters $(\mathcal{G}_i, \beta)$. If our graph is merely $\mathcal{G}$, we write $\mathbb{P}_\beta$. In other situations, if we consider coupled random variables with different parameters, we omit subscripts and write $\mathbb{P}$. Finally, for a statistic S, we may write S_i to denote that the statistic is compute with respect to $\mathcal{G}_i$, or for a graph $\mathcal{G}$, we may write $S_\mathcal{G}$ to emphasize the computation with respect to $\mathcal{G}$.

3. Maximum Likelihood and Monotonicity

In this chapter, we consider two approaches to inference that might be found in an introductory course: maximum likelihood and monotonicity. The first refers to the usual maximum likelihood estimate of parameters, where here the parameters are both the underlying graph $\mathcal{G}$ and scalar β. The second is the monotonicity of a p-value or decision threshold as we vary the underlying parameter β. To simplify matters, we shall only consider the case of no censoring, or $c = 0$, for this chapter.

3.1 Maximum Likelihood

A first standard approach is to consider maximum lielihood estimation. The good news is that some standard maximum likelihood results are sufficiently general to apply in some sense. However, it is the more standard setting of maximum likelihood estimation in which we are given m i.i.d. samples from the same distribution, which would require multiple infections in this case. We let $L(\mathcal{G}, \beta; J)$ denote the likelihood. These results follow from Section 3.3 of Lehmann and Romano (2006).

Lemma 3.1.1. *Suppose that the distributions SSM$(\mathcal{G}, \beta)$ are distinct, i.e. the parameters are identifiable. Suppose we receive m i.i.d. infections $\mathcal{I}^{(1)}, \ldots, \mathcal{I}^{(m)} \sim$ SSM$(\mathcal{G}_0, \beta_0)$. For any fixed $(\mathcal{G}, \beta) \neq (\mathcal{G}_0, \beta_0)$ as $m \to \infty$, we have*

$$\mathbb{P}_{0,\beta_0}\Big(L\Big(\mathcal{G}_0, \beta_0; \mathcal{I}^{(1:m)}\Big) > L\Big(\mathcal{G}, \beta; \mathcal{I}^{(1:m)}\Big)\Big) \to 1.$$

The proof is identical to that of Lehmann and Romano (2006), and so we omit it. Before stating the proposition, we define the likelihood equation as

$$\frac{\partial}{\partial \beta} \log L\Big(\mathcal{G}, \beta; \mathcal{I}^{(1:m)}\Big).$$

Thus a root $\hat{\beta}$ of the likelihood equation is just a parameter for which this derivative is 0.

Proposition 3.1.1. *Let β_0 be contained in $(0, \infty)$. Under the same conditions as the previous lemma, there is a sequence $\hat{\theta}_m = \Big(\hat{\mathcal{G}}_m, \hat{\beta}_m\Big)$ where (i) $\hat{\mathcal{G}}_m$ maximizes the likelihood across graphs, (ii) $\hat{\beta}_m$ are roots of the likelihood equation, and (iii) $\hat{\theta}_m$ tends to the true parameter value $\theta_0 = (\mathcal{G}_0, \beta_0)$ in probability.*

Note that the condition that β_0 be in the open interval from 0 to ∞ is so that we can characterize the estimate $\hat{\beta}_m$ as a zero of the derivative. We also require $\beta > 0$, since allowing

$\beta = 0$ makes the graph unidentifiable. A slight modification of the proof would extend the result to the empty graph. Another reason to prefer $\beta < \infty$ is that it makes it easier to avoid non-identifiability issues. For example, consider the case of a line graph with $n = 8$ vertices where $k = 5$ are infected. Then, if $\beta = \infty$, the fourth and fifth vertices are always infected, making the underlying graph unidentifiable, even as m increases. If $\beta < \infty$, then any two vertices can have different infection statuses

Of course, if it has not become obvious yet, the real problem with maximum likelihood estimation is computational. Since our stochastic spreading model requires summing probabilities over paths, in general we would have to sum over $k!$ paths in order to compute the likelihood for a single infection. Supposing that we wanted to approximate the likelihood, we could try sampling infection paths as follows. Given a single infection J, let $U(\mathcal{G}, \beta; J)$ be the random variable defined by

$$U(\mathcal{G}, \beta; J) = \begin{cases} \mathbb{P}_{\mathcal{G},\beta}\{\mathcal{P} = P\} & \text{w.p. } \frac{1}{k!} \text{ for } P \in \mathfrak{P}(J) \\ 0 & \text{otherwise.} \end{cases}$$

Then, we have

$$\mathbb{E}[U(\mathcal{G}, \beta; J)] = \frac{1}{k!} \sum_{P \in \mathfrak{P}(J)} \mathbb{P}_{\mathcal{G},\beta}\{\mathcal{P} = U\} = \frac{1}{k!} L(\mathcal{G}, \beta; J).$$

If we consider m uniform samples $U_1, \ldots, U_m$, then we can obtain an unbiased estimate of the likelihood by averaging and scaling, yielding

$$\tilde{L}_m(\mathcal{G}, \beta; J) := \frac{k!}{m} \sum_{i=1}^{m} U_i(\mathcal{G}, \beta; J).$$

We could define a similar random variable on the log scale if desired. Of course, the major problem is still the $k!$ term, because the variance of $\tilde{L}_m(\mathcal{G}, \beta; J)$ may be very high even for moderately large m.

3.2 Monotonicity

Consider the following basic statistical hypothesis test:

$$\begin{aligned} H_0 : &\quad X \sim \mathcal{N}\left(\mu_0, \sigma^2\right) \quad \mu_0 \leq 0 \\ H_1 : &\quad X \sim \mathcal{N}\left(\mu_1, \sigma^2\right) \quad \mu_1 > 0 \end{aligned}$$

where σ^2 is a known constant. Then, if we consider $\mu_0 = 0$, we might reject the null hypothesis when the Z-statistic

$$Z = \frac{X}{\sqrt{\sigma^2}}$$

is sufficiently large. In particular, we reject when Z exceeds the threshold t_α, which is given by

$$t_\alpha = \sup\{z \in \text{supp}(Z) : \mathbb{P}_{\mu_0=0}(Z \geq z) > \alpha\}.$$

Of course, we determine a p-value for the test numerically by using the fact that $Z \sim \mathcal{N}(0,1)$. Now, what if we want to construct a valid test for general $\mu_0' \leq 0$? If we define a statistic Z' by

$$Z' = \frac{X - \mu_0'}{\sqrt{\sigma^2}} = Z + C,$$

where C is non-negative. Then, by examining thresholds, we have

$$t_\alpha' = \sup\Big\{ z \in \text{supp}(Z') : \mathbb{P}_{\mu_0 = \mu_0'}(Z' \geq z) > \alpha \Big\}.$$

Since Z is a standard normal random variable when $\mu_0 = 0$ and Z' is a standard normal random variable when $\mu_0 = \mu_0' < 0$, we observe that the thresholds $t_\alpha = t_\alpha'$. However, in the latter case, we only require $Z + C = Z' > t_\alpha$ to reject the null hypothesis, and this is always satisfied when $Z > t_\alpha$. Put another way, we have

$$t_\alpha = \sup_{\mu_0 \leq 0} \sup\{ z \in \text{supp}(Z) : \mathbb{P}_{\mu_0}(Z \geq z) > \alpha \} = \sup\{ z \in \text{supp}(Z) : \mathbb{P}_{\mu_0 = 0}(Z \geq z) > \alpha \} \tag{3.1}$$

In terms of p-values, the p-value obtained for any $\mu_0' < 0$ is strictly smaller than that of the p value for $\mu_0 = 0$. This is a natural monotonicity that simplifies the testing problem from both a mathematical and computational point of view.

With a clearer idea of what we mean by monotonicity, we consider it in the problem of graph testing. Suppose we have the following testing problem:

$$\begin{aligned} H_0 : &\quad \mathcal{I} \sim \text{SSM}(\mathcal{G}_0, \beta_0) \quad \beta_0 \in \mathfrak{B}_0 \\ H_1 : &\quad \mathcal{I} \sim \text{SSM}(\mathcal{G}_1, \beta_1) \quad \beta_1 \in \mathfrak{B}_1 \end{aligned}$$

for some parameter set $\mathfrak{B}_0$ and $\mathfrak{B}_1$. Suppose that we use a statistic $S(\mathcal{I})$, and we reject the null hypothesis when $S(\mathcal{I})$ is greater than some threshold t_α. To control the Type I error at a level α, we need to have

$$t_\alpha = \sup_{\beta_0 \in \mathfrak{B}_0} \max\{ s \in \text{supp}(S) : \mathbb{P}_{\mathcal{G}_0, \beta_0}(S(\mathcal{I}) \geq s) > \alpha \}. \tag{3.2}$$

Unlike equation (3.1), it is far from clear that there is any natural monotonicity that we can use to remove the maximum over the parameter in computing this threshold.

Our first hope might be that the threshold for a simple statistic such as the edges within $\mathcal{G}_0$ is monotonically increasing in β_0. Indeed, some of our simulation results, this seems to be the case on the very small number of values of β_0 we examine. Unfortunately, this turns out to not be the case in general. For simplicity, let $t_\alpha(\beta_0)$ be the threshold value with respect to a single parameter value β_0

Proposition 3.2.1. *Consider the hypothesis test*

$$\begin{aligned} H_0 : &\quad \mathcal{I} \sim \mathit{SSM}(\mathcal{G}_0, \beta_0) \quad \beta_0 \in [0, \infty] \\ H_1 : &\quad \mathcal{I} \sim \mathit{SSM}(\mathcal{G}_1, \beta_1) \quad \beta_1 \in \mathfrak{B}_1 \end{aligned}$$

where $\mathcal{G}_0 \neq \mathcal{G}_1$. Fix any level α, and reject the null hypothesis when W_0 is sufficiently large. Then, for sufficiently large n and k, there is a graph $\mathcal{G}_0$ such that $t_\alpha(0) > t_\alpha(\infty)$.

To keep the exposition separate from the proofs, we prove this result in Section 3.4. However, the graph $\mathcal{G}_0$ is very simple to describe. It consists of two connected components, one of which is large but very sparse and the other of which is small but fully connected.

If we were somehow able to rule out such pathological cases with massive disparities in connectivity, perhaps one could prove a monotonicity result. We conjecture that there is a class of graphs for which this is true.

Conjecture 3.2.1. *For a fixed α and $k > 0$, if $\mathcal{G}_0$ is a vertex-transitive graph, then the threshold $t_\alpha(\beta_0)$ for the statistic W_0 is monotonically increasing in β_0.*

3.2.1 Monotonicity in the Ising Model

While we are mainly interested in our stochastic spreading model, we do note that the desried monotonicity does hold for the Ising model.

Proposition 3.2.2. *Consider the hypothesis test*

$$\begin{aligned} H_0 : &\quad \mathcal{I} \sim \mathit{Ising}(\mathcal{G}_0, \beta_0) \quad \beta_0 \in \mathfrak{B}_0 \\ H_1 : &\quad \mathcal{I} \sim \mathit{Ising}(\mathcal{G}_1, \beta_1) \quad \beta_1 \in \mathfrak{B}_1. \end{aligned}$$

Fix any level α, and reject the null hypothesis when W_0 is sufficiently large. Then, the threshold t_α is monotonically increasing as a function of β.

3.3 Discussion

In this chapter, we discussed maximum likelihood estimation and monotonicity of the thresholds. Both of these techniques have shortcomings that make them insufficient for practical use by themselves, and this is one of the motivations for our discretization and permutation methods. Note that we shall revisit maximum likelihood in the coming chapters, particularly as it relates to discretization and to the permutation statistics.

3.4 Proofs

This section contains proofs of the results of this chapter.

Proof of Proposition 3.1.1. First, we examine the average log-likelihood, which we define as

$$f\Big(\mathcal{G}, \beta; \mathcal{I}^{(1:m)}\Big) = \frac{1}{m} \sum_{i=1}^{m} \log L\Big(\mathcal{G}, \beta; \mathcal{I}^{(i)}\Big).$$

Our first goal is to show that this function is suitably well-behaved. Taking the derivative

with respect to β, we have

$$\begin{aligned}
\frac{\partial}{\partial\beta} f\Big(\mathcal{G},\beta;\mathcal{I}^{(1:m)}\Big) &= \frac{1}{m}\sum_{i=1}^{m}\left(\sum_{P\in\mathfrak{P}(\mathbb{I}_{k,0})}\prod_{t=1}^{k}\frac{1+\beta W_{t,v}(P)}{n+1-t+\beta W_t(P)}\right)^{-1} \\
&\quad\times\sum_{P\in\mathfrak{P}(\mathbb{I}_{k,0})}\sum_{t=1}^{k}\left(\prod_{s\neq t}\frac{1+\beta W_{s,v}(P)}{n+1-t+\beta W_s(P)}\right) \\
&\quad\times\frac{1}{n+1-t+\beta W_t(P)}\cdot\frac{W_{t,v}(P)(n+1-t)-W_t(P)}{n+1-t+\beta W_t(P)} \\
&= \frac{1}{m}\sum_{i=1}^{m}\sum_{P\in\mathfrak{P}(\mathbb{I}_{k,0})}\sum_{t=1}^{k}\frac{n+1-t+\beta W_t(P)}{1+\beta W_{t,v}(P)} \\
&\quad\times\frac{1}{n+1-t+\beta W_t(P)}\cdot\frac{W_{t,v}(P)(n+1-t)-W_t(P)}{n+1-t+\beta W_t(P)} \\
&= \frac{1}{m}\sum_{i=1}^{m}\sum_{P\in\mathfrak{P}(\mathbb{I}_{k,0})}\sum_{t=1}^{k}\frac{1}{1+\beta W_{t,v}(P)}\cdot\frac{W_{t,v}(P)(n+1-t)-W_t(P)}{n+1-t+\beta W_t(P)}.
\end{aligned}$$

Note that this is bounded above by

$$\frac{\partial}{\partial\beta} f\Big(\mathcal{G},\beta;\mathcal{I}^{(1:m)}\Big) \leq \frac{1}{m}\sum_{i=1}^{m}\sum_{P\in\mathfrak{P}(\mathbb{I}_{k,0})}\sum_{t=1}^{k}\frac{1}{1+\beta W_{t,v}(P)}\cdot\frac{W_{t,v}(P)(n+1-t)}{n+1-t+\beta W_t(P)},$$

so, the derivative has an upper bound of the form

$$\frac{\partial}{\partial\beta} f\Big(\mathcal{G},\beta;\mathcal{I}^{(1:m)}\Big) \leq \frac{a}{b+c\beta^2}$$

for constants $a\geq 0$ and $b,c>0$ that can be chosen independently of m. Thus, on an interval $[\beta,\beta']$ the average log-likelihood is bounded as

$$\begin{aligned}
f\Big(\mathcal{G},\beta'';\mathcal{I}^{(1:m)}\Big) &\leq f\Big(\mathcal{G},\beta;\mathcal{I}^{(1:m)}\Big) + \int_{\beta}^{\beta'}\frac{a}{b+cx^2}dx \\
&= f\Big(\mathcal{G},\beta;\mathcal{I}^{(1:m)}\Big) + \frac{a}{\sqrt{bc}}\left(\arctan\left(\beta'\sqrt{\frac{c}{b}}\right)-\arctan\left(\beta\sqrt{\frac{c}{b}}\right)\right).
\end{aligned}$$

With this representation, we see that if we pick a sufficiently fine finite discretization $\mathfrak{B}_D$ of $[0,\infty]$, we can approximate the average likelihood everywhere within an additive factor ϵ. Next, we want to pick ϵ sufficiently small to distinguish the graph $\mathcal{G}_0$ correctly.

Let Δ be defined by

$$\Delta = \mathbb{E}_{\mathcal{G}_0,\beta_0}\log L(\mathcal{G}_0,\beta_0;\mathcal{I}) - \sup_{\mathcal{G}\neq\mathcal{G}_0,\beta\in[0,\infty]}\mathbb{E}_{\mathcal{G}_0,\beta_0}\log L(\mathcal{G},\beta;\mathcal{I}).$$

By Jensen's inequality, we see that

$$\Delta > \inf_{\mathcal{G}\neq\mathcal{G}_0,\beta\in[0,\infty]}\log\mathbb{E}_{\mathcal{G}_0,\beta_0}\frac{L(\mathcal{G}_0,\beta_0;\mathcal{I})}{L(\mathcal{G},\beta;\mathcal{I})}\geq 0,$$

with strict inequality since the logarithm is strictly concave. Note that by the law of large numbers,

$$f\Big(\mathcal{G}, \beta; \mathcal{I}^{(1:m)}\Big) \xrightarrow{\text{a.s.}} \mathbb{E}_{\mathcal{G}_0,\beta_0} \log L(\mathcal{G}, \beta; \mathcal{I}),$$

so, we have

$$\mathbb{P}\bigg\{ f\Big(\mathcal{G}_0, \beta_0; \mathcal{I}^{(1:m)}\Big) > f\Big(\mathcal{G}, \beta; \mathcal{I}^{(1:m)}\Big) + \frac{\Delta}{2} \bigg\} \to 1.$$

At this point, we pick $\epsilon = \Delta/2$, and we define the following event:

$$\begin{aligned} E_m(\delta) = \Big\{ \mathcal{I}^{(1:m)} \in \mathbb{I}_{k,0}^m : & f\Big(\mathcal{G}_0, \beta_0; \mathcal{I}^{(1:m)}\Big) > f\Big(\mathcal{G}_0, \beta_0 - \delta; \mathcal{I}^{(1:m)}\Big) \\ & f\Big(\mathcal{G}_0, \beta_0; \mathcal{I}^{(1:m)}\Big) > f\Big(\mathcal{G}_0, \beta_0 + \delta; \mathcal{I}^{(1:m)}\Big) \\ & f\Big(\mathcal{G}_0, \beta_0; \mathcal{I}^{(1:m)}\Big) > f\Big(\mathcal{G}, \beta; \mathcal{I}^{(1:m)}\Big) + \epsilon \\ & \text{for all } \mathcal{G} \neq \mathcal{G}_0, \beta \in \mathfrak{B}_D \}. \end{aligned}$$

By Lemma 3.1.1 and a union bound, we see that for any fixed δ, we have

$$\mathbb{P}_{0,\beta_0}(E_m(\delta)) \to 1.$$

By a diagonal argument, we can pick a sequence $\delta_m \to 0$ such that

$$\mathbb{P}_{0,\beta_0}(E_m(\delta_m)) \to 1$$

as well.

Now, we can define $\hat{\theta}_m = \big(\hat{\mathcal{G}}, \hat{\beta}_m\big)$. We let $\hat{\mathcal{G}}_m = \mathcal{G}_0$ since this maximizes the likelihood across graphs, and we let $\hat{\beta}_m$ to be the zero of the derivative of the log-likelihood with respect to β under the event $E_m(\delta_m)$ in the interval $(\beta_0 - \delta_m, \beta_0 + \delta_m)$. Then we observe the probability that $\hat{\beta}_m$ is a root of the likelihood equation is

$$\mathbb{P}_{0,\beta_0}\bigg(\frac{\partial f}{\partial \beta}\Big(\hat{\mathcal{G}}_m, \hat{\beta}_m; \mathcal{I}^{(1:m)}\Big)\bigg) \geq \mathbb{P}_{0,\beta_0}(E_m(\delta_m)) \to 1.$$

Finally, we observe that the parameter estimates converge to the true values. Let $\delta > 0$ be an arbitrary constant, and if m is sufficiently large, we have

$$\begin{aligned} \mathbb{P}_{0,\beta_0}\Big(\hat{\mathcal{G}}_m = \mathcal{G}_0,\ |\hat{\beta}_m - \beta_0| \leq \delta\Big) &\geq \mathbb{P}_{0,\beta_0}\Big(\hat{\mathcal{G}}_m = \mathcal{G}_0,\ |\hat{\beta}_m - \beta_0| \leq \delta_m\Big) \\ &= \mathbb{P}_{0,\beta_0}(E_m(\delta_m)) \to 1, \end{aligned}$$

and this completes the proof. □

Proof of Proposition 3.2.1. Define a graph $\mathcal{G}$ to be the union of two graphs $\mathcal{G}'$ and $\mathcal{G}''$. We let $\mathcal{G}'$ be a cycle on $n' = (1 - \alpha + \gamma)n$ vertices, and let $\mathcal{G}''$ be a complete graph on $n'' = (\alpha - \gamma)n$ vertices for appropriate $\gamma > 0$. It suffices to consider the cases of $\beta = 0$ and $\beta = \infty$.

In the former case, let X be the number of infected vertices in $\mathcal{G}'$. Then, we can write

$$\mathbb{P}(X = k') = \frac{\binom{n'}{k'}\binom{n''}{k''}}{\binom{n}{k}}.$$

Thus, X has a Hypergeometric(n', n, k) distribution. By Theorem 1 and Theorem 4 of Hoeffding (1963), we have the inequality

$$\mathbb{P}(X \geq (1 - \alpha + \gamma + t)k) = \mathbb{P}\left(X \geq \left(\frac{n'}{n} + t\right)k\right) \leq \exp\left(-2t^2 k\right). \tag{3.3}$$

So for any fixed $t > 0$, this probability can be made arbitrarily small by increasing k.

Let A denote the event on the left hand side of equation (3.3), and let $Y = k - X$ be the number of infected vertices in $\mathcal{G}''$. Then under event A, we see $Y \leq (\alpha - \gamma - t)k$. Thus, we can lower bound the threshold $t_\alpha(0)$ for the edges within statistic by

$$(\alpha - \gamma - t)^2 k^2 \leq t_\alpha(0)$$

assuming that the right hand side of equation (3.3) is upper bounded by α, i.e. when

$$k \geq \frac{1}{2t^2} \log \frac{1}{\alpha}. \tag{3.4}$$

Next, we consider the case where $\beta = \infty$. Then, with probability $1 - (\alpha - \gamma)$, the resulting infection forms a connected subgraph in $\mathcal{G}'$, and so $W(\mathcal{I}) = k - 1$. Thus, we are guaranteed

$$t_\alpha(\infty) \leq k \leq (\alpha - \gamma - t)^2 k^2 \leq t_\alpha(0)$$

when

$$\frac{1}{(\alpha - \gamma - t)^2} \leq k. \tag{3.5}$$

For fixed α, equation (3.4) and equation (3.5) can be satisfied by taking n and k to be sufficiently large, completing the proof. □

Proof of Proposition 3.2.2. For a fixed β_0, let A be the set

$$A = \{J : W(J) > t_\alpha(\beta_0)\}.$$

Then, we can define

$$f(\beta) = \frac{\sum_{J \in A} \exp(\beta W(J))}{\sum_{J \in \mathbb{I}_{k,0}} \exp(\beta W(J))}.$$

It suffices to show that the derivative of f is positive, since this would imply that the threshold t_α is monotonically increasing in β. Differentiating, we have

$$\begin{aligned} f'(\beta) &= \frac{\sum_{J \in A} W(J) \exp(\beta W(J))}{\sum_{J \in \mathbb{I}_{k,0}} \exp(\beta W(J))} - \frac{\sum_{J \in A} \exp(\beta W(J))}{\sum_{J \in \mathbb{I}_{k,0}} \exp(\beta W(J))} \cdot \frac{\sum_{J \in \mathbb{I}_{k,0}} W(J) \exp(\beta W(J))}{\sum_{J \in \mathbb{I}_{k,0}} \exp(\beta W(J))} \\ &= \frac{\sum_{J \in A} W(J) \exp(\beta W(J))}{\sum_{J \in \mathbb{I}_{k,0}} \exp(\beta W(J))} - f(\beta) \frac{\sum_{J \in \mathbb{I}_{k,0}} W(J) \exp(\beta W(J))}{\sum_{J \in \mathbb{I}_{k,0}} \exp(\beta W(J))}. \end{aligned}$$

At this point, we would like to observe that this is non-negative as follows. Define the sequence $a_1 \geq \ldots \geq a_m$ to be the values of $W(J)$ in decreasing order, and let $b_1 \geq \ldots \geq b_m$ be the values of $\exp(\beta W(J))/\sum_{J \in \mathbb{I}_{k,0}} \exp(\beta W(J))$ in decreasing order. Then, it suffices to show that

$$\sum_{i=1}^{|A|} a_i b_i \geq f(\beta) \sum_{i=1}^{m} a_i b_i.$$

If we let $b'_i = b_i / f(\beta)$, note that both the b_i and b'_i sum to 1, so we need to show that

$$\sum_{i=1}^{|A|} a_i b'_i \geq \sum_{i=1}^{m} a_i b_i.$$

Now, we define $g : [0,1] \to \mathbb{R}$ and $h : [0,1] \to \mathbb{R}$ in the following manner. We let

$$g(t) = \sum_{i=1}^{j-1} a_i b'_i + a_j \left(t - \sum_{i=1}^{j-1} b'_i \right) \quad \text{where} \quad \sum_{i=1}^{j} b'_i \leq t \leq \sum_{i=1}^{j+1} b'_i$$

$$h(t) = \sum_{i=1}^{j-1} a_i b_i + a_j \left(t - \sum_{i=1}^{j-1} b_i \right) \quad \text{where} \quad \sum_{i=1}^{j} b_i \leq t \leq \sum_{i=1}^{j+1} b_i.$$

Now, we simply need to show that $g(1) \geq h(1)$. But, we observe that $g(0) = h(0) = 0$, and $g'(t) \geq h'(t)$ almost everywhere. Thus, $g(1) \geq h(1)$, which completes the proof. $\square$

Part II

Discretization

4. Theory

In this chapter, we examine the possibility of discretizing our null hypothesis space. Through a discretization, we could hope to approximately compute a threshold t_α, even with a supremum over parameter values as in equation (3.2). Our approach here is more general, as we use bounds on total variation distance. The main tool is a data-processing inequality for f-divergences, but we defer the proof details to Chapter 5.

4.1 Main Result

In this section, we provide our main results. The overall idea is straightforward. We have a space of probability measures $\mathcal{M}(\mathcal{G}, \mathfrak{C})$. We wish to find a δ-net in total variation distance for the set of distributions $\mathcal{M}(\mathcal{G}, \mathfrak{B}) \subseteq \mathcal{M}(\mathcal{G}, \mathfrak{C})$, and we do this by providing the appropriate finite subset $\mathfrak{B}_D \subset \mathfrak{B}$. Using this finite subset, we can approximately simulate the possible null distributions to conduct hypothesis tests and build confidence intervals. Thus, the main technical details of the paper consist of finding the discretization $\mathfrak{B}_D$.

A natural question is how to construct a discretization $\mathfrak{B}_D$ for a given $\mathfrak{B}$. Our strategy is to first construct a finite set $\mathfrak{C}_D$ of $\mathfrak{C}$ such that $\mathcal{M}(\mathcal{G}, \mathfrak{C}_D)$ yields a δ-net of $\mathcal{M}(\mathcal{G}, \mathfrak{C})$. Then, we can define a discretization function $F_D : \mathfrak{C} \to \mathfrak{C}_D$ such that

$$\mathrm{TV}\Big(\mathbb{P}_\beta, \mathbb{P}_{F_D(\beta)}\Big) \leq \delta, \tag{4.1}$$

where TV is the total variation distance. Finally, for any $\mathfrak{B} \subseteq C$, we can construct a discretization $\mathfrak{B}_D := F_D(\mathfrak{B})$.

The first question one might have is whether such a $\mathfrak{C}_D$ exists. While our construction proves that it does, we can also reason that this must be true from a topological perspective. Using the standard topology associated with $\mathfrak{C}$ and the topology on $\mathcal{M}(\mathcal{G}, \mathfrak{C})$ induced by the total variation distance as a metric, the compactness of $\mathfrak{C}$ and the continuity of the function $\mathcal{M}(\mathcal{G}, \cdot)$ imply that $\mathfrak{M}$ must also be compact, which is certainly sufficient for the existence of a δ-net.

Given the existence of a cover $\mathfrak{C}_D$ one might wonder why bother with $\mathfrak{B}_D = F(\mathfrak{B})$ at all because, in order to obtain a valid test, we could simply use $\mathfrak{C}_D$. From a statistical perspective, if we would like a more powerful test, then it is beneficial to reduce the size of the δ-blowup of $\mathcal{M}(\mathcal{G}, \mathfrak{B}_D)$, which in this case involves eliminating superfluous points of $\mathfrak{C}_D$ when forming $\mathfrak{B}_D$. In particular, it is important to eliminate small values of β if possible, since such β produce measures $\mathcal{M}(\mathcal{G}, \beta) = \mathbb{P}_\beta$ that do not depend strongly on $\mathcal{G}$. From a computational viewpoint, eliminating superfluous β allows us to conduct our tests with less computation.

In the remaining subsections, we introduce our discretization, give a simulation-based algorithm to control the Type I error at a level α, and offer the coresponding $(1 - \alpha)$-confidence set procedure.

4.1.1 Discretization

We consider the following discretization with its associated discretization function.

Definition 4.1.1. *Let $\mathcal{G}$ be a graph. Let*

$$M_0 = \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \frac{W_t(P_{1:t-1})}{n+1-t}$$

Define M_1 to be an integer satisfying

$$M_1 \geq \frac{1}{2\delta} M_0.$$

Define M_2 to be an integer such that

$$M_2 \geq \frac{k+c}{2\delta}.$$

Let N_0 be an integer such that

$$N_0 \geq (k+c)\left(n - \frac{k+c-1}{2}\right)$$

Let N_1 be an integer satisfying

$$N_1 \geq \frac{k+c}{2\delta}\left(\log \frac{N_0}{\delta} + \log \frac{M_1}{M_2}\right).$$

We define the one-dimensional discretization $\mathfrak{C}_D$ to include the following values of β:

- $\beta_i = \frac{i}{M_1}$, *for* $i = 0, 1, \ldots, M_2$, *and*
- $\beta_{M_2+i} = \beta_{M_2+i-1} \exp\left(\frac{2\delta}{k+c}\right)$ *for* $i = 1, \ldots, N_1$.

Finally, the one-dimensional discretization function F_D associated with this discretization is

$$F_D(\beta) := \begin{cases} \arg\min_{\beta' \in \mathfrak{C}_D} |\beta - \beta'| & \beta \leq \frac{M_2}{M_1} \\ \max\left\{\beta' : \beta' \leq \beta \leq \exp\left(\frac{\delta}{k+c}\right)\beta'\right\} & \frac{M_2}{M_1} \leq \beta \leq \frac{N_0}{\delta} \\ \min\left\{\beta' : \exp\left(-\frac{\delta}{k+c}\right)\beta' < \beta \leq \beta'\right\} & \frac{M_2}{M_1} \leq \beta \leq \frac{N_0}{\delta} \\ \max\{\beta' \in \mathfrak{C}_D\} & \frac{N_0}{\delta} < \beta. \end{cases}$$

Note that for β satisfying $M_2/M_1 \leq \beta \leq N_0/\delta$, the β' is unique by the definition of the discretization. For the graph $\mathcal{G}_i$, we shall simply denote this set by $\mathfrak{C}_{i,D}$. Additionally, we denote the size of the set by $N = |\mathfrak{C}_D| = M_2 + N_1 + 1$. Now, we have the following proposition.

Proposition 4.1.1. *Let $\mathcal{G}$ be a graph. Then, the discretization $\mathfrak{C}_D$ with discretization function F_D satisfies equation (4.1).*

Algorithm 4.1.1: One-Statistic Test

Input : Type I error tolerance $\alpha > 0$, approximation parameters ϵ, δ, γ, and ξ, observed infection vector $\mathcal{I}$, null graph $\mathcal{G}_0$, statistic S, discretization $\mathfrak{B}_{0,D}$, discretization function F_D

1 Define $N_{\text{sims}} \geq \left(\frac{1}{2\epsilon^2} + \frac{8}{3\epsilon}\right) \log \frac{N}{\xi}$.

2 For each β in $\mathfrak{B}_{0,D}$, simulate an infection N_{sims} times on $\mathcal{G}_0$ to obtain the approximate $1 - (\alpha - \gamma - \epsilon)$ quantile $\hat{t}_{\alpha-\gamma-\epsilon,\beta}$.

3 Compute $\hat{t}_{\alpha-\gamma-\epsilon} = \max_{\beta \in \mathfrak{B}_{0,D}} \hat{t}_{\alpha-\gamma-\epsilon,\beta}$.

4 Reject the null hypothesis if $S(\mathcal{I}) > \hat{t}_{\alpha-\gamma-\epsilon}$.

4.1.2 Hypothesis Testing Algorithm

With a discretization satisfying equation (4.1) in hand, it is straightforward to devise a simulation-based algorithm that controls the Type I error. Thus, we present Algorithm 4.1.1.

Theorem 4.1.1. *Let $\mathcal{G}_0$ be a graph and S be a statistic. Let $\alpha > 0$ be given, and set $\gamma = \delta + \xi$. Set $\epsilon < \alpha - \gamma$. Then, Algorithm 4.1.1 with a discretization and discretization function satisfying equation (4.1) controls the Type I error at a level α.*

We prove this in Section 5.1. While Theorem 4.1.1 proves the correctness of simulation tests, we still need to address further statistical and computational details. Statistically, we are also interested in tests with good power. Since the test statistic S may be arbitrary, we cannot make statements about the power of the procedure in general. For example, a constant statistic S would have a Type I error of 0 and a Type II error of 1. Nonetheless, this does allow us to use a broad class of statistics. This does not include the likelihood ratio, even though the likelihood ratio suffers from the further problem of being generally uncomputable. We discuss approximations to the likelihood ratio in the following section.

4.1.3 Confidence Set Algorithm

When discussing hypothesis tests, a natural thing to do is to invert the tests to obtain confidence sets. Since we do not know the structure of these sets, particularly if they form connected intervals, we refrain from referring to these sets as "confidence intervals." To this end, we provide a slightly modified algorithm to compute confidence sets, matching the test that we have given.

To do this, we again want to use our discretization and discretization function. Now in analog with Algorithm 4.1.1, we provide the following confidence set algorithm.

Corollary 4.1.1. *If the discretization $\mathfrak{B}_D$ and the discretization function F_D satisfy equation (4.1), then the set C returned by Algorithm 4.1.2 is a $(1-\alpha)$-confidence set.*

Algorithm 4.1.2: One-Statistic Confidence Set

Input : Type I error tolerance $\alpha > 0$, approximation parameters ϵ, δ, γ, and ξ, observed infection vector $\mathcal{I}$, graph $\mathcal{G}$, statistic S, discretization $\mathfrak{B}_D$, discretization function F_D

1 Define $N_{\text{sims}} \geq \left(\frac{1}{2\epsilon^2} + \frac{8}{3\epsilon}\right) \log \frac{N}{\xi}$.

2 For each β in $\mathfrak{B}_D$, simulate an infection N_{sims} times on $\mathcal{G}$ to obtain the approximate $1 - (\alpha - \gamma - \epsilon)$ quantile $\hat{t}_{\alpha-\gamma-\epsilon,\beta}$.

3 Define the discrete confidence set

$$C_D := \left\{\beta \in \mathfrak{B}_D : S(\mathcal{I}) \leq \hat{t}_{\alpha-\gamma-\epsilon,\beta}\right\}.$$

4 Return the confidence set $C = F_D^{-1}(C_D)$.

4.1.4 Computation

Now, we comment on the computational aspects of this algorithm. To run this algorithm, we first need to compute M_0. Each summand yields a cardinality-constrained weighted max-cut problem. In general, the associated unconstrained decision problem is NP-complete (Karp, 1972), but there are polynomial-time approximation algorithms (Goemans and Williamson, 1995) and polynomial time algorithms for planar graphs (Hadlock, 1975). Alternatively, we can use the sum of the largest $t-1$ weighted degrees to bound each $W_t(P)$. A good, computable choice is the minimum of this crude degree bound and a scaled arpproximate solution of the maxcut problem.

Next, we need to run $N \times N_{\text{sims}}$ simulations. Using a simple weighted degree upper bound, we can give an upper bound on the number of simulations required. Let D be the maximum weighted degree in the graph. Then, it suffices to have

$$N \geq 1 + \frac{k+c}{2\delta}\left(1 + \log\frac{N_0}{\delta} + \log\frac{D}{k+c} \cdot \sum_{t=1}^{k+c} \frac{t}{n+1-t}\right).$$

Fortunately, this only depends logarithmically on the size of the graph n. Additionally, since the order of the simulations does not matter, each of them can be run in parallel. However, even for graphs of moderate size, this may be too many simulations for testing the entirety of $\mathfrak{C}$. Thus, it may be necessary to restrict to a small set $\mathfrak{B}$. If we wish to restrict ourselves to the range $[0, R]$ for $R > M_2/M_1$, then one could replace the N_0/δ in the logarithm above with R to bound the number of necessary simulations.

Finally, we make a brief remark on choosing δ, ϵ, and ξ. All of these make the approximate threshold worse, so from a statistic perspective, we would like to minimize each of these. From a computational perspective, we would like all of these to be large. This tradeoff means that we have to select how to allocate a statistical approximation budget between the three or, alternatively, decide how small we can make each of these based on computational resources. The important point is that while all affect the statistical approximation of the threshold equally, they do not affect the computation equally. As a rough approximation of

$N \times N_{\text{sims}}$, the dependence on δ is δ^{-1}; the dependence on ϵ is ϵ^{-2}; and the dependence on ξ is $\log \xi^{-1}$. Thus, a sensible computational approach may be to make these terms roughly equal, i.e. have $\delta^{-1} \approx \epsilon^{-2} \approx \log \xi^{-1}$.

4.2 Extensions

In this section, we discuss three extensions. The first of these is an extension to likelihood-based test statistics. Second, we discuss multidimensional weights. Finally, we make a few remarks on graph confidence sets Proofs for all of the extensions are given in Chapter 5.

4.2.1 Likelihood-Based Tests

The setup that we have discussed so far has let us consider statistics such as $W_1 - cW_0$ for some constant c. However, the most obvious statistic to use is the log-likelihood ratio. Specifically, the log-likelihood ratio is

$$\ell(\beta_0, \beta_1; J) = \log L(\mathcal{G}_1, \beta_1; J) - \log L(\mathcal{G}_0, \beta_0; J) = \ell_1(\beta_1; J) - \ell_0(\beta_0; J).$$

Since the log-likelihood ratio is difficult to compute for our stochastic spreading model, we might also be interested in approximations

$$S(\beta_0, \beta_1; J) = S_1(\beta_1; J) - S_0(\beta_0; J), \tag{4.2}$$

where S_i is a computable approximation of ℓ_i. In particular, we saw in Chapter 3 that

$$S_W(\beta_0, \beta_1; J) = \beta_1 W_1(I) - \beta_0 W_0(I)$$

is the first-order approximation of the log-likelihood ratio when $c = 0$.

However, there is a key difference in this case compared to the statistics considered in the preceding section. Here, S is a function of $\mathcal{I}$ *and* the parameters β_0 and β_1. So, there are two choices that we consider to overcome this difficulty. First, instead of finding a threshold $\hat{t}$ to compare $S(I)$ against, we find a function $\hat{t} : \mathfrak{B}_0 \times \mathfrak{B}_1 \to \mathbb{R}$ such that $\hat{t}(\beta_0, \beta_1) \leq S(I; \beta_0, \beta_1)$ with probability at most α under the null hypothesis across values of β_0 and β_1. The second option is to skirt dependencies on the parameter value in the statistic. For example, using S_W as inspiration, we might consider a class of linear statistics

$$S(\beta_0, \beta_1; J) = \beta_1 S_1(I) - \beta_0 S_0(I). \tag{4.3}$$

Then by considering S_0 and S_1 separately, we can reject the null hypothesis if for each value of β_0, either S_0 is too small or S_1 is too large.

Since the threshold function method is more cumbersome, we defer it to Chapter 5. Now, we consider the second approach: two-statistic tests. Let $s_{0,1-\alpha,\beta}$ and $s_{1,\alpha,\beta}$ denote the α and $1-\alpha$ quantiles of S_0 and S_1 respectively, and let $\hat{s}_{0,1-\alpha,\beta}$ and $\hat{s}_{1,\alpha,\beta}$ denote their empirical counterparts. Thus, we can state Algorithm 4.2.1.

Proposition 4.2.1. *If the discretization $\mathfrak{B}_D$ with discretization function F_D satisfies equation (4.1), then Algorithm 4.2.1 controls the Type I error at a level α.*

Algorithm 4.2.1: Two-Statistic Test

Input : Type I error tolerances α_0 and α_1 where $\alpha_0 + \alpha_1 = \alpha$, approximation parameters ϵ, δ, γ, and ξ, observed infection vector $\mathcal{I}$, null graph $\mathcal{G}_0$, statistics S_0 and S_1, discretization $\mathfrak{B}_{0,D}$, discretization function F_D

1 Define $N_{\text{sims}} \geq \left(\frac{1}{2\epsilon^2} + \frac{8}{3\epsilon}\right) \log \frac{N}{\xi}$.

2 For each β in $\mathfrak{B}_{0,D}$, simulate an infection N_{sims} times on $\mathcal{G}_0$ to obtain the approximate the $(\alpha_0 - \gamma - \epsilon)$ quantile $\hat{s}_{0,1-\alpha_0-\gamma-\epsilon,\beta}$ and the $1 - (\alpha_1 - \gamma - \epsilon)$ quantile $\hat{s}_{1,(\alpha_1-\gamma-\epsilon),\beta}$.

3 Reject the null hypothesis if for each $\beta \in \mathfrak{B}_{0,D}$, either $S_0 < \hat{s}_{0,1-(\alpha_0-\gamma-\epsilon),\beta}$ or $S_1 > \hat{s}_{1,(\alpha_1-\gamma-\epsilon),\beta}$.

Now, Proposition 4.2.1 is hardly revolutionary in terms of dividing the α budget between two separate tests. However, in a simple example, this is shown to lead to a better Type II error than other methods, and we imagine that this is helpful even in cases where the Type II error is difficult to compute exactly.

4.2.2 Multidimensional Weights

In this section, we consider a modification of our previous setup. Here, the edge weights $\boldsymbol{w}(u, v) = (w^1(u, v), \ldots, w^d(u, v))$ are vectors in $\mathbb{R}^d_+$. As a result, we have $\mathfrak{C} = \mathbb{R}^d_+$, and so we denote our parameter vectors by $\boldsymbol{\beta}$. In our stochastic spreading model, the exponential random variables T_{uv} now have parameter $\boldsymbol{\beta}^\intercal \boldsymbol{w}(u, v)$. Additionally, in this context $W_{t,v}$ and W_t are also vectors in $\mathbb{R}^d_+$. The former is the multidimensional infection weight into vertex v at time t, and the latter is the multidimensional weighted cut between the infected and uninfected vertices at time t. We denote the ith coordinate of these by $W^i_{t,v}$ and W^i_t respectively. Additionally, we refer to the edges where $w^i(u, v) > 0$ as being edges of type i.

However, there is one additional difficulty in the multidimensional case that does not appear in one dimension. In one dimension, we have let $\beta = \infty$ to enforce spreading over edges when there is an edge between an infected and an uninfected vertex. From a topological view, this is a natural way to obtain a compact $\mathfrak{C}$ that also maintains the continuity of the parametrization $\mathcal{M}$.

In the multidimensional case, enforcing spreading over edges depends on the relative importance of the different dimensions of the edge weights. Thus, finding a compactification of $\mathfrak{C} = \mathbb{R}^d_+$ that preserves the continuity of the parametrization $\mathcal{M}$ is simply far more difficult for $d > 1$. We give an illustration of this in Figure 4.1. As a result, we shall stick to a discretization for bounded $\mathfrak{B}$ for $d > 2$ and for a potentially unbounded parameter set when $d = 2$. Now, we can define the discretization.

Definition 4.2.1. *Let $\mathcal{G}$ be a graph, and let $\mathfrak{C}_R = [0, R]^d$. Suppose the edge weights are in $\mathbb{R}^d_+$. Let M^i_0 be*

$$M^i_0 = d(d + 1) \sum_{t=1}^{k+c} \frac{W^i_t(P_{1:t-1})}{n + 1 - t}$$

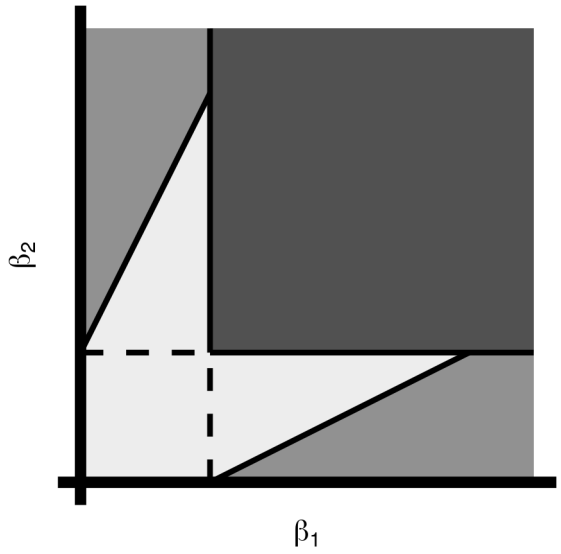

Figure 4.1 Discretization regions for $d = 2$. The lightly shaded hexagon is discretized by $\mathfrak{C}_{R,D}$. The dark square is discretized by Δ_D^2. Finally, the top left and bottom right trapezoidal regions are discretized by $\mathfrak{H}_{1,D}$ and $\mathfrak{H}_{2,D}$ respectively.

Define M_1^i to be an integer satisfying

$$M_1^i \geq \frac{1}{2\delta} M_0^i.$$

Let M_2^i be an integer such that

$$M_2^i \geq \frac{d(d+1)(k+c)}{2\delta}$$

Let N_1^i be an integer satisfying

$$N_1^i \geq \left(\log_{1+\frac{2\delta}{d(d+1)(k+c)}} R + \log_{1+\frac{2\delta}{d(d+1)(k+c)}} \frac{M_1^i}{M_2^i} \right).$$

Define a one-dimensional discretization $\mathfrak{C}_{R,D,i}$ to include the following values of β:

- $\beta_j = \frac{j}{M_1^i}$ *for* $j = 0, 1, \ldots, M_2^i$,
- $\beta_{M_2^i+j} = \beta_{M_2^i+j-1}\left(1 + \frac{2\delta}{d(d+1)(k+c)}\right)$ *for* $j = 1, \ldots, N_1^i$.

Then, the multidimensional discretization is $\mathfrak{C}_{R,D} = \mathfrak{C}_{R,D,1} \times \cdots \times \mathfrak{C}_{R,D,d}$, i.e. the cartesian product of d-univariate discretizations. Finally, the discretization function is

$$F_{R,D}(\boldsymbol{\beta}) := (F_{R,D,1}(\beta_1), \ldots, F_{R,D,d}(\beta_d))$$

where

$$F_{R,D,i}(\beta) = \max\{\beta' \in \mathfrak{C}_{D,R,i} : \beta' \leq \beta\}.$$

Lemma 4.2.1. *The set $\mathfrak{C}_{R,D}$ with discretization function $F_{R,D}$ satisfies equation (4.1).*

Since Algorithm 4.1.1 and Algorithm 4.2.1 only depend on having a discretization satisfying equation (4.1), we can use them for multidimensional testing.

Now, we consider the 2-dimensional case with large $\boldsymbol{\beta}$. There are two kinds of large $\boldsymbol{\beta}$: the case in which both β_1 and β_2 are large and another in which only one of the two is large. From a topological perspective, to compactify the upper right quadrant and maintain the continuity of parametrization function, we need to add three line segments at infinity, two of which are almost mirror images of each other.

We start with the former. To differentiate this case, in place of $\boldsymbol{\beta}$, we use the parameter $\boldsymbol{\eta}$ instead, and our process evolves according to

$$\mathbb{P}_\eta(\mathcal{P}_t = v|\mathcal{P}_{1:t-1} = P_{1:t-1}) = \begin{cases} \frac{W_{t,v}(P_{1:t-1})^\intercal \boldsymbol{\eta}}{W_t(P_{1:t-1})^\intercal \boldsymbol{\eta}} & W_t(P_{1:t-1}) \neq \mathbf{0} \\ \frac{1}{n+1-t} & W_t(P_{1:t-1}) = \mathbf{0}, v \notin P_{1:t-1} \\ 0 & v \in P_{1:t-1}. \end{cases}$$

Additionally, we only need to consider $\boldsymbol{\eta}$ in the 2-simplex

$$\Delta^2 := \Big\{(\eta_1, \eta_2) \in [0,1]^2 : \eta_1 + \eta_2 = 1\Big\}.$$

At this point, we can consider the discretization.

Definition 4.2.2. *Let $\mathcal{G}$ be a given graph with a weight function w. Then, let M_∞ be an integer satisfying*

$$M_\infty \geq \min_{i=1,2} \max_{P \in \mathfrak{P}} \sum_{t=1}^{k+c} W_t^i(\mathcal{P}_{1:t-1})$$

Then, we define the discretization

$$\Delta_D^2 := \left\{\left(\frac{i}{M_\infty}, \frac{M_\infty - i}{M_\infty}\right) : i = 0, \ldots, M_\infty\right\}.$$

Finally, we define the discretization function

$$F_{\Delta,D}(\boldsymbol{\beta}) = \operatorname*{arg\,min}_{(\eta_1, 1-\eta_1) \in \Delta_D^2} \left|\eta_1 - \frac{\beta_1}{\beta_1 + \beta_2}\right|$$

where

$$\beta_1, \beta_2 \geq \frac{2n(k+c)}{\delta}.$$

Lemma 4.2.2. *Let $\mathcal{G}$ be a graph where all non-zero edge weights are grater than 1, i.e. for $(u,v) \in \mathcal{E}$, $w_i(u,v) > 0$ for an $i = 1, \ldots, d$ implies $w_i(u,v) \geq 1$. The discretization Δ_D^2 with discretization function $F_{\Delta,D}$ satisfies equation (4.1).*

Now, we consider the second type of of large $\boldsymbol{\beta}$. Without loss of generality, assume that β_1 is small and β_2 is large. In this case, we use a modified spreading process $\mathcal{S}$. Here, to choose the infected vertex $\mathcal{S}_t$ at time t, we spread over an edge of type two if we can do so, and if we cannot, then the procedure evolves according to our univariate spreading model with parameter ζ and edge weights w^1, i.e. the edges of the first type. Now, we define the necessary discretization.

Definition 4.2.3. *Let $\mathcal{G}$ be a given graph. Let M_0^i be*

$$M_0^i = \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \frac{W_t^i(P_{1:t-1})}{n+1-t}.$$

Define M_1^i to be an integer satisfying

$$M_1^i \geq \frac{1}{2\delta} M_0^i.$$

Define M_2^i to be an integer such that

$$M_2^i \geq \frac{2(k+c)}{\delta}$$

Let N_0 be

$$N_0 = (k+c)\left(n - \frac{k+c-1}{2}\right)$$

Let N_1 be an integer satisfying

$$N_1 \geq 2\left(\frac{k+c}{\delta}\right)\left(\log\frac{N_0}{\delta} + \log\frac{M_1^i}{M_2^i}\right).$$

Then, we define $\mathfrak{H}_{i,D}$ to consist of the following values of ζ:

- $\zeta_j = \frac{j}{M_1^i}$ *for* $j = 0, 1, \ldots, M_2^i$,
- $\zeta_{M_2^i+j} = \zeta_{M_i^\star+j-1} \exp\left(\frac{\delta}{2(k+c)}\right)$ *for* $i = 1, \ldots, N_1$.

Define the constants

$$K = \frac{2n(k+c)}{\delta}$$

and

$$Q = \frac{M_0^i}{\delta}$$

in forming $\mathfrak{H}_{i,D}$. Finally, the discretization function is

$$F_{\mathfrak{H},i,D}(\boldsymbol{\beta}) = \max\{\zeta \in \mathfrak{C}_{D,R} : \zeta \leq \beta_i\}$$

where $F_{\mathfrak{H},i,D}$ is only defined for $\boldsymbol{\beta}$ satisfying $K + Q\beta_i \leq \beta_j$ for $j \neq i$.

Lemma 4.2.3. *The discretization $\mathfrak{H}_{i,D}$ with discretization function $F_{\mathfrak{H},i,D}$ satisfies equation (4.1).*

Finally, we want to conclude that the three discretizations together comprise a discretization of $\mathfrak{C} = \mathbb{R}^2_+$. To this end, we need a discretization function of the entire space. For simplicity, let $R = K = 2n(k+c)/\delta$, and let Q be defined as in Definition 4.2.3. Then, we define

$$F_{D,2}(\boldsymbol{\beta}) := \begin{cases} F_{\Delta,D}(\boldsymbol{\beta}) & R \leq \beta_1, \beta_2 \\ F_{\mathfrak{H},1,D}(\boldsymbol{\beta}) & \beta_1 \leq R \text{ and } K + Q\beta_1 \leq \beta_2 \\ F_{\mathfrak{H},2,D}(\boldsymbol{\beta}) & \beta_2 \leq R \text{ and } K + Q\beta_2 \leq \beta_1 \\ F_{R,D}(\boldsymbol{\beta}) & \text{otherwise.} \end{cases}$$

Now, we can state the corollary.

Corollary 4.2.1. *Let $R = K(1+Q)$ where K and Q are defined in Definition 4.2.3. The discretization $\mathfrak{C}_D = \mathfrak{C}_{R,D} \cup \Delta^2_D \cup \mathfrak{H}_{1,D} \cup \mathfrak{H}_{2,D}$ with discretization function $F_{D,2}$ satisfies equation (4.1).*

4.2.3 Graph Confidence Sets

While Algorithm 4.1.2 and its two-statistic equivalent are phrased as a algorithms for constructing confidence sets for the parameter β, note that we can use these algorithms to construct confidence sets of pairs $(\mathcal{G}, \beta)$. For example, we might be interested in knowing not only if our HIV graph is in an $(1-\alpha)$-confidence set for a given range of β, but we might also wish to know whether close variants of the graph are also in the confidence set.

One curious detail is the choice of statistic when testing different graphs. Our theory allows for different statistics for each graph, and so for graph $\mathcal{G}$, it might be reasonable to use a statistic such as $W_{\mathcal{G}}$. If we make this choice though, note that if we consider $\mathcal{G}$ to be the empty or complete graphs, then the statistic is constant and so $(\mathcal{G}, \beta)$ is in the confidence set for any β and any confidence level. We can think of this as a specific case of centering the confidence set. If we are fairly certain that our confidence set should include a particular graph $\mathcal{G}^\star$, then for testing a graph $\mathcal{G}$, we can use the statistic such $S_{\mathcal{G}} = W_{\mathcal{G}} - W_{\mathcal{G}^\star}$. Then, we see that $S_{\mathcal{G}^\star}$ is constant, which is the phenomenon we observed earlier for the empty graph.

It is not entirely clear why this occurs or if there is a good alternative. We could specify a fixed statistic S, which would not lead to a constant statistic on any graph but may not capture the structure of the graph being tested. Of course, we could always consider the intersection of two $(1-\alpha/2)$-confidence sets, but this does not answer the question of how to choose the statistics for each of the confidence sets.

4.3 An Example

In this section, we analyze risk bounds in a simple case.

We start by considering $\mathcal{G}_0$ to be an $(n-1)$-star and $\mathcal{G}_1$ to be the n-cycle. Without loss of generality, let vertex 1 be the center of the star so that $\mathcal{I}_1$ denotes the status of the center. We shall specify n and k later. For simplicity, we set $\alpha_0 = \alpha_1 = \alpha/2$ and use the actual thresholds $s_{0,1-\alpha/2}$ and $s_{1,\alpha/2}$ instead of the estimated thresholds since we can approximate these arbitrarily closely by conducting more simulations. In order to demonstrate power over pre-existing methods, we need to restrict the parameter sets. In this case, we consider $\mathfrak{B}_0 = [1, \infty]$ and $\mathfrak{B}_1 = [1, \infty]$.

The statistics we shall use are $S_i(\mathcal{I}) = W_i(\mathcal{I})$. From Proposition 7.3.3, we see that

$$\mathbb{P}_{0,\beta_0}(W_0(\mathcal{I}) = 0) \leq \exp\left(-\frac{k + \beta_0 k(k-1)/2}{n - k + 1 + \beta_0(k-1)}\right) \leq \exp\left(-\frac{k(k-1)}{2n}\right).$$

For this to be less than $\alpha/2$, we require that

$$n \log \frac{2}{\alpha} \leq \frac{k(k-1)}{2}.$$

Thus, a Type I error is committed when $W_0(\mathcal{I}) = 0$ or when $W_1(\mathcal{I})$ is sufficiently small.

Now, we can compute the probability of a Type II error. First, we note that under $\mathbb{P}_{1,\beta_1}$, the events $\{\mathcal{I}_1 = 1\}$ and $\{W_1(\mathcal{I}) = w\}$ are independent. We can see this by computing

$$\begin{aligned}
\mathbb{P}_{1,\beta_1}(\{\mathcal{I}_1 = 1\} \cap \{W_1(\mathcal{I}) = w\}) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}_{1,\beta_1}(\{\mathcal{I}_i = 1\} \cap \{W_1(\mathcal{I}) = w\}) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{1,\beta_1}[\mathbf{1}\{\mathcal{I}_i = 1\}\mathbf{1}\{W_1(\mathcal{I}) = w\}] \\
&= \frac{1}{n}\mathbb{E}_{1,\beta_1}\left[\left(\sum_{i=1}^{n} \mathbf{1}\{\mathcal{I}_i = 1\}\right)\mathbf{1}\{W_1(\mathcal{I}) = w\}\right] \\
&= \frac{1}{n}\mathbb{E}_{1,\beta_1}[k\mathbf{1}\{W_1(\mathcal{I}) = w\}] \\
&= \frac{k}{n}\mathbb{P}_{0,\beta_1}(W_1(\mathcal{I}) = w),
\end{aligned}$$

where the first equality follows from the exchangeability of the vertices. Since we can easily see that the events $\{W_0(\mathcal{I}) = k-1\}$ and $\{\mathcal{I}_1 = 1\}$ are equivalent, the Type II error is

$$(\text{TII}) = \mathbb{P}_{1,\beta_1}\Big(\{\mathcal{I}_1 = 1\} \cap \{W_1(\mathcal{I}) \geq s_{1,\alpha/2}\}\Big) = \frac{k}{n}\mathbb{P}_{1,\beta_1}\Big(W_1(\mathcal{I}) \geq s_{1,\alpha/2}\Big)$$

Using Corollary 8.1.2, we see can bound this final probability so that

$$(\text{TII}) \leq \frac{k}{n}\exp\left(-\frac{1}{2k}\left((k-1)2^{k-1}\prod_{m=1}^{k-1}\frac{\beta_1}{n-m+2\beta_1} - \frac{k(k-1)}{n-1} - \sqrt{2k\log\frac{2}{\alpha}}\right)^2\right)$$

Note that this gives a superior Type II error over the permutation method of Chapter 7 by k/n at the cost of a factor of 2 inside of the logarithm, which is essentially a constant term. However, we needed to restict the possible parameter values, and the resulting computation is far more time-consuming.

4.4 Discussion

In this chapter, we provided methods for testing infection spread between any two graphs $\mathcal{G}_0$ and $\mathcal{G}_1$ and the spaces of parameter values $\mathfrak{B}_0$ and $\mathfrak{B}_1$ under the stochastic spreading model. This is a very general method that allows arbitrary graphs to be tested in practice.

There are still a variety of questions that remain unresolved. Do tests with substantially smaller discretizations exist? Our method requires N points. Is this optimal in general, or are there discretizations that are much smaller? While it may not be possible to improve substantially on the total variation distance argument, perhaps monotonicity of the type considered in Chapter 3 could reduce the size of the discretizations in special cases.

5. Technical Details

In this chapter, we provide proofs of our results for discretization.

5.1 Proofs

In this section, we provide the main discretization and algorithm proofs.

5.1.1 Discretization Proof

The purpose of this section is to prove Proposition 4.1.1. To assist, we start with the following proposition.

Proposition 5.1.1. *Let $\underline{\beta} = \min\{\beta, \beta'\}$, and let $\overline{\beta} = \max\{\beta, \beta'\}$. For finite β and β', we have the bound*

$$TV(\mathbb{P}_\beta, \mathbb{P}_{\beta'}) \leq \min\left\{|\beta - \beta'|M_0,\ \frac{k+c}{2}\log\frac{\overline{\beta}}{\underline{\beta}},\ \frac{N_0}{\underline{\beta}}\right\}.$$

Proof. Let $\mathcal{I}$ and $\mathcal{I}'$ denote infections according to $\mathbb{P}_\beta$ and $\mathbb{P}_{\beta'}$ respectively. First, we want a simpler representation of $\mathcal{I}$ and $\mathcal{I}'$. Let C be c censored random variables, which may be chosen in any manner, such as uniformly random choices or a fixed set of vertices, as long as this choice is independent of the path random variables to be defined now. Let $\mathcal{P}$ be path random variable of length $k+c$ for parameter β. To generate this, let X_t denote the indicator of a spread over the edges at time t, i.e.

$$\mathbb{P}(X_t = 1 | \mathcal{P}_{1:t-1} = P_{1:t-1}) = \frac{W_t(P_{1:t-1})\beta}{n+1-t+W_t(P_{1:t-1})\beta}.$$

Next, we denote Y_t to be a uniform choice a random vertex, i.e.

$$\mathbb{P}(Y_t = v | \mathcal{P}_{1:t-1} = P_{1:t-1}) = \begin{cases} \frac{1}{n+1-t} & v \notin \mathcal{P}_{1:t-1} \\ 0 & v \in \mathcal{P}_{1:t-1}. \end{cases}$$

Further, if $X_t = 0$, then $\mathcal{P}_t = Y_t$. Finally, we denote Z_t to be a random vertex chosen according to

$$\mathbb{P}(Z_t = v | \mathcal{P}_{1:t-1} = P_{1:t-1}) = \frac{W_{t,v}(P_{1:t-1})}{W_t(P_{1:t-1})},$$

and if $X_t = 1$, then we set $\mathcal{P}_t = Z_t$. Then, we can write

$$\mathcal{I} = f(C, \mathcal{P}) = g(C, X_{1:k+c}, Y_{1:k+c}, Z_{1:k+c}),$$

where the functions f and g may be defined as follows. For f, the vertices of C are censored. Then, the first k uninfected variables of $\mathcal{P}$ are chosen to be infected, and the remaining

choices are ignored. The function g is chosen similarly, since we are simply breaking the path into its various choices. Note that this gives the desired measure on $\mathcal{I}$. Similarly, we can write

$$\mathcal{I}' = f(C, \mathcal{P}') = g(C, X'_{1:k+c}, Y'_{1:k+c}, Z'_{1:k+c}),$$

for analogous choices X'_t, Y'_t, and Z'_t for the parameter β' and path $\mathcal{P}'$. Further, we couple $\mathcal{P}$ and $\mathcal{P}'$ such that if $\mathcal{P}_{1:t-1} = \mathcal{P}'_{1:t-1}$, then $Y_t = Y'_t$ and $Z_t = Z'_t$. Additionally, assume that X_t and X'_t are maximally coupled.

Let $\mathbb{P}_X$ denote the measure with respect to a random variable X. Now, the first step is to use the data-processing inequality of Lemma 5.5.1 and Lemma 5.5.2 to obtain

$$\begin{aligned}
\mathrm{TV}(\mathbb{P}_\beta, \mathbb{P}_{\beta'}) &\le \mathrm{TV}\Big(\mathbb{P}_{C, X_{1:k+c}, Y_{k+c}, Z_{k+c}}, \mathbb{P}_{C, X'_{1:k+c}, Y'_{k+c}, Z'_{k+c}}\Big) \\
&\le \mathbb{P}\big((X'_{1:k+c}, Y'_{k+c}, Z'_{k+c}) \neq (X'_{1:k+c}, Y'_{k+c}, Z'_{k+c})\big) \\
&\le \mathbb{P}\big(X_{1:k+c} \neq X'_{1:k+c}\big),
\end{aligned}$$

where the final inequality follows due to the coupling.

Now, we define the set $\mathfrak{P}_t$ to be

$$\mathfrak{P}_t := \{(X_{1:t}, Y_{1:t}, Z_{1:t}, X'_{1:t}, Y'_{1:t}, Z'_{1:t}) : X_{1:t} = X'_{1:t}\}.$$

Then, we have

$$\begin{aligned}
\mathrm{TV}(\mathbb{P}_\beta, \mathbb{P}_{\beta'}) &\le \sum_{t=1}^{k+c} \mathbb{P}(X_{1:t-1} = X'_{1:t-1}) \mathbb{P}(X_t \neq X'_t | X_{1:t-1} = X'_{1:t-1}) \\
&\le \sum_{t=1}^{k+c} \sum_{E \in \mathfrak{P}_{t-1}} \mathbb{P}(E) \mathbb{P}(X_t \neq X'_t | E) \qquad (5.1) \\
&\le \sum_{t=1}^{k+c} \max_{E \in \mathfrak{P}_{t-1}} \mathbb{P}(X_t \neq X'_t | E).
\end{aligned}$$

Equation (5.1) yields

$$\mathrm{TV}(\mathbb{P}_\beta, \mathbb{P}_{\beta'}) \le \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \left| \frac{W_t(P_{1:t-1})\beta}{n+1-t+W_t(P_{1:t-1})\beta} - \frac{W_t(P_{1:t-1})\beta'}{n+1-t+W_t(P_{1:t-1})\beta'} \right|,$$

where this follows from the maximal coupling of X_t and X'_t, i.e. a Bernoulli B_p of parameter p and a Bernoulli B_q of parameter q are maximally coupled if $\mathrm{TV}(\mathbb{P}_{B_p}, \mathbb{P}_{B_q}) = |p - q|$.

The next step is to examine the function

$$h(x) = \frac{ax}{b + ax}$$

where $a, b \ge 0$. The derivative of h is

$$h'(x) = \frac{a}{b+ax}\left(1 - \frac{ax}{b+ax}\right) = \frac{a}{b+ax}\left(\frac{b}{b+ax}\right) \le \min\left\{\frac{a}{b}, \frac{1}{2x}, \frac{b}{ax^2}\right\}.$$

We apply this to our present problem to obtain

$$\begin{aligned}
\mathrm{TV}(\mathbb{P}_\beta, \mathbb{P}_{\beta'}) &\leq \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \int_{\underline{\beta}}^{\overline{\beta}} \min\left\{ \frac{W_t(P_{1:t-1})}{n+1-t},\ \frac{1}{2x},\ \frac{n+1-t}{W_t(P_{1:t-1})x^2} \right\} dx \\
&\leq \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \min\left\{ \int_{\underline{\beta}}^{\overline{\beta}} \frac{W_t(P_{1:t-1})}{n+1-t} dx,\ \int_{\underline{\beta}}^{\overline{\beta}} \frac{1}{2x} dx,\ \int_{\underline{\beta}}^{\overline{\beta}} \frac{n+1-t}{W_t(P_{1:t-1})x^2} dx \right\} \\
&= \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \min\left\{ |\beta-\beta'| \frac{W_t(P_{1:t-1})}{n+1-t},\ \frac{1}{2}\log\frac{\overline{\beta}}{\underline{\beta}},\ \frac{n+1-t}{W_t(P_{1:t-1})}\left(\frac{1}{\underline{\beta}} - \frac{1}{\overline{\beta}}\right) \right\} \\
&\leq \min\left\{ |\beta-\beta'| \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \frac{W_t(P_{1:t-1})}{n+1-t},\ \frac{k+c}{2}\log\frac{\overline{\beta}}{\underline{\beta}},\ \frac{1}{\underline{\beta}} \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \frac{n+1-t}{W_t(P_{1:t-1})} \right\} \\
&\leq \min\left\{ |\beta-\beta'| M_0,\ \frac{k+c}{2}\log\frac{\overline{\beta}}{\underline{\beta}},\ \frac{k+c}{\underline{\beta}}\left(n - \frac{k+c-1}{2}\right) \right\} \\
&= \min\left\{ |\beta-\beta'| M_0,\ \frac{k+c}{2}\log\frac{\overline{\beta}}{\underline{\beta}},\ \frac{N_0}{\underline{\beta}} \right\},
\end{aligned}$$

and this completes the proof. □

At this point, we can finally prove our main proposition.

Proof of Proposition 4.1.1. Let β be an element of $\mathfrak{C}$, and let $\mathcal{I}$ be an infection generated on $\mathcal{G}$ with parameter β. Now, we need to show that $\mathfrak{C}_D$ has an element β' such that an infection $\mathcal{I}'$ on $\mathcal{G}$ with parameter β' has a distribution close to that of $\mathcal{I}$. We consider the cases where β is in $[0, M_2/M_1]$, $[M_2/M_1, N_0/\delta]$, and $[N_0/\delta, \infty]$.

First, suppose β is in $[0, M_2/M_1]$. Then, $\beta' = F_D(\beta)$ satisfies $|\beta - \beta'| \leq 1/(2M)$. By Proposition 5.1.1, we have

$$\mathrm{TV}(\mathbb{P}_\beta, \mathbb{P}_{\beta'}) \leq |\beta-\beta'| M_0 \leq \frac{M_0}{2M_1} \leq \delta.$$

which is what we wanted to show.

Next, we consider the case of β in $[M_2/M_1, N_0/\delta]$. Let $\beta' = F_D(\beta)$. Then, we have

$$\overline{\beta} \leq \exp\left(\frac{2\delta}{k+c}\right)\underline{\beta}.$$

Using this and Proposition 5.1.1, we have

$$\mathrm{TV}(\mathbb{P}_\beta, \mathbb{P}_{\beta'}) \leq \frac{k+c}{2}\log\frac{\overline{\beta}}{\underline{\beta}} \leq \delta.$$

Thus, we once again see that β and β' satisfy equation (4.1).

Finally, we consider the case of $\beta \geq N_0/\delta$. Then, we have $\beta' = F_D(\beta) \geq N_0/\delta$, and by Proposition 5.1.1, we have

$$\mathrm{TV}(\mathbb{P}_\beta, \mathbb{P}_\infty) \leq \frac{N_0}{\beta} \leq \delta.$$

This completes the proof of the proposition. □

5.1.2 Algorithm Proofs

Our main goal is to prove Theorem 4.1.1. This this end, we separate the Type I error into two components: the "actual" Type I error for an unusual observation and a Type I error resulting from simulation inaccuracies. We start with a lemma to this effect.

Lemma 5.1.1. *Let t be a real number, let $\hat{t}$ be a random threshold independent of $\mathcal{I}$, and let $S : \mathbb{I}_{k,c} \to \mathbb{R}$ be a statistic. Then, we have*

$$\mathbb{P}_{0,\beta}\Big(S(\mathcal{I}) \geq \hat{t}\Big) \leq \mathbb{P}_{0,\beta}(S(\mathcal{I}) \geq t) + \mathbb{P}_{0,\beta}\Big(\hat{t} \leq t\Big).$$

The proof is a series of straightforward manipulations that we give in Section 5.2. Next, we want to use basic concentration results to show that quantiles of statistics concentrate sufficiently nicely.

Lemma 5.1.2. *Let $S : \mathbb{I}_{k,c} \to \mathbb{R}$ be a statistic. If*

$$N_{sims} \geq \left(\frac{1}{2\epsilon^2} + \frac{2}{3\epsilon}\right) \log \frac{1}{\xi},$$

then we have

$$\mathbb{P}_{0,\beta}\Big(\hat{t}_{\alpha-\epsilon,\beta} < t_\alpha\Big) \leq \xi.$$

The proof of this lemma is also given in Section 5.2 Finally, we can prove Theorem 4.1.1.

Proof of Theorem 4.1.1. First, we use Proposition 4.1.1, Lemma 5.1.1, and equation (2.1) to obtain

$$\begin{aligned}
\sup_{\beta \in \mathfrak{B}_0} \mathbb{P}_{0,\beta}\Big(S(\mathcal{I}) \geq \hat{t}_{\alpha-\gamma-\epsilon}\Big) &\leq \sup_{\beta \in \mathfrak{B}_{0,D}} \mathbb{P}_{0,\beta}\Big(S(\mathcal{I}) > \hat{t}_{\alpha-\gamma-\epsilon}\Big) + \delta \\
&\leq \sup_{\beta \in \mathfrak{B}_{0,D}} \mathbb{P}_{0,\beta}(S(\mathcal{I}) \geq t_{\alpha-\gamma,\beta}) + \mathbb{P}_{0,\beta}\Big(\hat{t}_{\alpha-\gamma-\epsilon} < t_{\alpha-\gamma,\beta}\Big) + \delta \\
&\leq \alpha - \gamma + \delta + \sup_{\beta \in \mathfrak{B}_{0,D}} \mathbb{P}_{0,\beta}\Big(\hat{t}_{\alpha-\gamma-\epsilon,\beta} < t_{\alpha-\gamma,\beta}\Big).
\end{aligned}$$

Thus, it simply remains to analyze this supremum. Since the supremum of positive numbers is bounded by their sum, we have

$$\sup_{\beta\in\mathfrak{B}_{0,D}} \mathbb{P}_{0,\beta}\Big(\hat{t}_{\alpha-\gamma-\epsilon} < t_{\alpha-\gamma}\Big) \le \sum_{\beta\in\mathfrak{B}_{0,D}} \mathbb{P}_{0,\beta}\Big(\hat{t}_{\alpha-\gamma-\epsilon} < t_{\alpha-\gamma}\Big).$$

By Lemma 5.1.2, each summand is bounded by ξ/N. Thus, we have

$$\sup_{\beta\in\mathfrak{B}_0} \mathbb{P}_{0,\beta}\Big(S(\mathcal{I}) \ge \hat{t}_{\alpha-\gamma-\epsilon}\Big) \le \alpha-\gamma+\delta+\sum_{\beta\in\mathfrak{B}_{0,D}} \frac{\xi}{N} \quad = \alpha-\gamma+\delta+\xi \le \alpha,$$

and this completes the proof. □

5.2 Additional Proofs

In this section, we have proofs for the supporting lemmas used in Section 5.1.

Proof of Lemma 5.1.1. We use straightforward union bounds to obtain

$$\begin{aligned}
\mathbb{P}_{0,\beta}\Big(S(\mathcal{I}) \ge \hat{t}\Big) &\le \mathbb{P}_{0,\beta}\Big(\Big(\Big\{S(\mathcal{I}) \ge \hat{t}\Big\} \cap \Big\{\hat{t} \ge t\Big\}\Big) \cup \Big(\Big\{S(\mathcal{I}) \ge \hat{t}\Big\} \cap \Big\{\hat{t} \le t\Big\}\Big)\Big) \\
&\le \mathbb{P}_{0,\beta}\Big(\Big\{S(\mathcal{I}) \ge \hat{t}\Big\} \cap \Big\{\hat{t} \ge t\Big\}\Big) + \mathbb{P}_{0,\beta}\Big(\Big\{S(\mathcal{I}) \ge \hat{t}\Big\} \cap \Big\{\hat{t} \le t\Big\}\Big) \\
&\le \mathbb{P}_{0,\beta}(S(\mathcal{I}) \ge t) + \mathbb{P}_{0,\beta}\Big(\hat{t} \le t\Big).
\end{aligned}$$

This completes the proof. □

Proof of Lemma 5.1.2. The goal of the proof is to apply Bernstein's inequality, given as Lemma 5.5.5. First by equation (2.1), define

$$p = \mathbb{P}_{0,\beta}(S(\mathcal{I}_i) \ge t_\alpha) > \alpha.$$

Then, we have

$$\begin{aligned}
\mathbb{P}_{0,\beta}\Big(\hat{t}_{\alpha-\epsilon,\beta} < t_\alpha\Big) &= \mathbb{P}_{0,\beta}\left(\frac{1}{N_{\text{sims}}}\sum_{i=1}^{N_{\text{sims}}} \mathbf{1}\{S(\mathcal{I}_i) \ge t_\alpha\} \le \alpha-\epsilon\right) \\
&= \mathbb{P}_{0,\beta}\left(\frac{1}{N_{\text{sims}}}\sum_{i=1}^{N_{\text{sims}}} \mathbf{1}\{S(\mathcal{I}_i) \ge t_\alpha\} - p \le -\epsilon-(p-\alpha)\right).
\end{aligned}$$

Now, by considering $X_i = p - \mathbf{1}\{S(\mathcal{I}_i) \ge t_\alpha\}$, we have $\mathbb{E}[X_i] = 0$, $|X_i| \le 1$, and we also see $\mathbb{E}[X_i^2] = p(1-p) \le 1/4$. So, we can apply Bernstein's inequality to obtain

$$\mathbb{P}_{0,\beta}\Big(\hat{t}_{\alpha-\epsilon,\beta} < t_\alpha\Big) \le \exp\left(-\frac{N_{\text{sims}}(\epsilon+(p-\alpha))^2}{1/2+2(\epsilon+(p-\alpha))/3}\right).$$

Thus, in order for this to be less than or equal to δ, we require

$$N_{\text{sims}} \ge \left(\frac{1}{2(\epsilon+p-\alpha)^2} + \frac{2}{3(\epsilon+p-\alpha)}\right)\log\frac{1}{\xi} \ge \left(\frac{1}{2\epsilon^2}+\frac{2}{3\epsilon}\right)\log\frac{1}{\xi},$$

and this completes the proof. □

Remark 5.2.1. One thing to note from the proof of Lemma 5.1.2 is that if $p = \mathbb{P}_{0,\beta}(S(\mathcal{I}_i) \geq t_\alpha)$ is strictly greater than α, then using N_{sims} simulations actually leads to a tighter bound on the error probability than ξ. Alternatively, we could achieve the desired error probability with fewer simulations, but since p is unknown, we settle for a coarser upper bound on the number of simulations required.

5.3 Extensions Proofs

In this section, we provide results for our extensions. We start with the likelihood-inspired threshold function test, and we give both the algorithm and the necessary proofs. Next, we prove our results for two-statistic tests. Finally, we prove our results for multidimensional parameter sets.

5.3.1 Threshold Function Test

In this section we consider our threshold function test. We start by introducing the algorithm, and then we prove the validity of the test.

Algorithm

The main difficulty is computing an appropriate threshold function for all values of β_0 and β_1 for a statisic S of the form given in equation (4.2). A straightforward solution to this problem is to impose a Lipschitz assumption on $S_1(I;\cdot)$ and $S_0(I;\cdot)$, which leads to the Lipschitz continuity of the threshold. Let $\mathfrak{B}_{1,D}$ be any finite subset of $\mathfrak{B}_1$, and let $\mathfrak{B}_D = \mathfrak{B}_{0,D} \times \mathfrak{B}_{1,D}$. Define the (L_0, L_1)-Lipschitz-penalized empirical quantile

$$\tilde{t}_{\alpha,L_0,L_1}(\beta_0,\beta_1) = \hat{t}_{\alpha,\beta_0'}(\beta_0',\beta_1') + (L_0|\beta_0 - \beta_0'| + L_1|\beta_1 - \beta_1'|),$$

where $\boldsymbol{\beta}' = F_D(\boldsymbol{\beta}) = (F_{0,D}(\beta_0), F_{1,D}(\beta_1))$ for a discretization function $F : \mathfrak{B}_0 \times \mathfrak{B}_1 \to \mathfrak{B}_D$.

Algorithm 5.3.1: Parameter-Dependent Lipschitz Test

Input : Type I error tolerance $\alpha > 0$, approximation parameters ϵ, δ, γ, and ξ, observed infection vector $\mathcal{I}$, null graph $\mathcal{G}_0$, statistic S, discretization $\mathfrak{B}_D$, discretization function F_D

1 Define $N_{\text{sims}} \geq \left(\frac{1}{2\epsilon^2} + \frac{8}{3\epsilon}\right) \log \frac{|\mathfrak{B}_D|}{\xi}$.

2 For each (β_0, β_1) in $\mathfrak{B}_D$, simulate an infection N_{sims} times on $\mathcal{G}_0$ to obtain the approximate $1 - (\alpha - \gamma - \epsilon)$ quantile $\hat{t}_{\alpha-\gamma-\epsilon,\beta_0}$.

3 Compute the function $\tilde{t}_{\alpha,L_0,L_1}(\beta_0,\beta_1)$ with discretization function f.

4 Reject the null hypothesis if $S(\mathcal{I};\beta_0,\beta_1) > \tilde{t}_{\alpha-\gamma-\epsilon}(\beta_0,\beta_1)$ for all $(\beta_0,\beta_1) \in \mathfrak{B}_0 \times \mathfrak{B}_1$.

Proposition 5.3.1. *Let S have the form given by equation (4.2). Additionally, suppose that $S_1(I;\cdot)$ is L_1-Lipschitz and $S_0(I;\cdot)$ is L_0-Lipschitz with probability 1. If the discretization $\mathfrak{B}_{0,D}$ with discretization function $F_{0,D}$ satisfies equation (4.1), then, Algorithm 5.3.1 using $\tilde{t}_{\alpha,L_0,L_1}$ controls the Type I error at a level α.*

Note that in order to obtain a discretization and discretization function satisfying equation (4.1), it is sufficient to use the discretization that we introduced in Definition 4.1.1.

While this is a good first step toward using likelihood approximations, this result may be too conservative to be useful for certain graphs. The main problem is the use of almost-sure Lipschitz assumptions, where the almost-sure Lipschitz constants may be far larger than what is seen in typical infections. A natural idea is to relax the almost-sure Lipschitz constants into high-probability Lipschitz constants that can be approximated via simulation. Unfortunately, doing this with the approximation techniques we have developed thus far does not seem to be possible.

Proofs

The goal of this section is to prove Proposition 5.3.1. As usual, we start with some helping lemmas.

Lemma 5.3.1. *Let $S(I;\beta_0,\beta_1) = S_1(I,\beta_1) - S_0(I,\beta_0)$ where $S_0(I,\cdot)$ and $S_1(I,\cdot)$ are L_0-Lipschitz and L_1-Lipschitz respectively almost surely. Then, $t_{\alpha,\beta_0}(\beta_0',\beta_1')$ is L_1-Lipschitz in β_1' and L_0-Lipschitz in β_0'.*

Proof. Define the set of supporting infections $X(\beta_0,\beta_1)$ as

$$X(\beta_0,\beta_1) := \{I \in \mathbb{I}_{k,c} : S(I;\beta_0,\beta_1) > t_{\alpha,\beta_0}(\beta_0,\beta_1)\}.$$

Note that we have

$$t_{\alpha,\beta_0}(\beta_0,\beta_1) = \max\{S(I;\beta_0,\beta_1) : I \notin X(\beta_0,\beta_1)\}.$$

Thus, we have $\mathbb{P}_{0,\beta_0}(\mathcal{I} \in X(\beta_0,\beta_1)) \leq \alpha$, and for any $t \in \operatorname{supp}(S(\cdot;\beta_0,\beta_1))$ such that $t < t_{\alpha,\beta_0}(\beta_0,\beta_1)$, we have

$$\mathbb{P}_{0,\beta_0}(S(\mathcal{I};\beta_0,\beta_1) \geq t) \geq \alpha.$$

For simplicity, let $\Delta = L_0|\beta_0 - \beta_0'| + L_1|\beta_1 - \beta_1'|$. Next, we prove the upper and lower bounds on $t_{\alpha,\beta_0}(\beta_0,\beta_1)$ separately. Now, since the S_i are L_i-Lipschitz, we see that

$$S(I;\beta_0',\beta_1') \leq S(I;\beta_0,\beta_1) + \Delta$$

for each I in $X(\beta_0,\beta_1)$. Thus, the probability that

$$S(\mathcal{I};\beta_0',\beta_1') \geq t_{\alpha,\beta_0}(\beta_0,\beta_1) + \Delta$$

is at most α. Therefore, we have

$$t_{\alpha,\beta_0}(\beta_0',\beta_1') \leq t_{\alpha,\beta_0}(\beta_0,\beta_1) + \Delta.$$

Next, we prove the lower bound. Analogously, we see that

$$S(I;\beta_0',\beta_1') \geq S(I;\beta_0,\beta_1) - \Delta$$

for each I in $X(\beta_0, \beta_1)$. Thus, the probability that

$$S(\mathcal{I}; \beta_0', \beta_1') \geq t_{\alpha,\beta_0}(\beta_0, \beta_1) - \Delta$$

is at least α. Therefore, we have

$$t_{\alpha,\beta_0}(\beta_0', \beta_1') \geq t_{\alpha,\beta_0}(\beta_0, \beta_1) - \Delta.$$

Putting everything together, we have

$$\left|t_{\alpha,\beta_0}(\beta_0', \beta_1') - t_{\alpha,\beta_0}(\beta_0, \beta_1)\right| \leq \Delta,$$

which proves the desired Lipschitz result. □

Lemma 5.3.2. *Define the event*

$$E := \{\tilde{t}_{\alpha-\gamma-\epsilon}(\beta_0', \beta_1') < t_{\alpha-\gamma,\beta_0}(\beta_0', \beta_1')\},$$

and the event

$$E' := \left\{\hat{t}_{\alpha-\gamma-\epsilon}(\beta_0, \beta_1) < t_{\alpha-\gamma,\beta_0}(\beta_0, \beta_1)\right\}$$

for $\beta_0 \in \mathfrak{B}_{0,D}$, $\beta_1 \in \mathfrak{B}_{1,D}$, $\beta_0 \in \mathfrak{B}_0$, and $\beta_1 \in \mathfrak{B}_1$. Then, we have the inclusion $E \subseteq E'$.

Proof. Suppose that E, holds. As before, let $\Delta = L_0|\beta_0 - \beta_0'| + L_1|\beta_1 - \beta_1'|$. Then by the definition of $\tilde{t}$, the assumption that E occurs, and Lemma 5.3.1, we have

$$\hat{t}_{\alpha-\gamma-\epsilon}(\beta_0, \beta_1) + \Delta = \tilde{t}_{\alpha-\gamma-\epsilon}(\beta_0', \beta_1') < t_{\alpha-\gamma,\beta_0}(\beta_0', \beta_1') \leq t_{\alpha-\gamma,\beta_0}(\beta_0, \beta_1) + \Delta.$$

Subtracting Δ, we have

$$\hat{t}_{\alpha-\gamma-\epsilon}(\beta_0, \beta_1) < t_{\alpha-\gamma,\beta_0}(\beta_0, \beta_1),$$

which completes the proof. □

Proof of Proposition 5.3.1. The proof proceeds in essentially the same manner as Theorem 4.1.1. However, we do have to take two additional steps. The first is to decouple the β_0 in the measure and in the statistic S from each other so that we can use a total variation distance argument. To this end, we have

$$\begin{aligned}(\text{TI}) &:= \sup_{\beta_0 \in \mathfrak{B}_0, \beta_1 \in \mathfrak{B}_1} \mathbb{P}_{0,\beta_0}\left(S(\mathcal{I}; \beta_0, \beta_1) \geq \tilde{t}_{\alpha-\gamma-\epsilon}(\beta_0, \beta_1)\right) \\ &\leq \sup_{\beta_0 \in \mathfrak{B}_0, \beta_1 \in \mathfrak{B}_1} \sup_{\beta_0' \in B(\beta_0)} \mathbb{P}_{0,\beta_0}\left(S(\mathcal{I}; \beta_0', \beta_1) \geq \tilde{t}_{\alpha-\gamma-\epsilon}(\beta_0', \beta_1),\right)\end{aligned}$$

where $B(\beta_0) = F_{0,D}^{-1}(F_{0,D}(\beta_0))$ for the discretization function

$$F_D(\beta_0, \beta_1) = (F_{0,D}(\beta_0), F_{1,D}(\beta_1)).$$

This is the necessary decoupling. Now, as in the proof of Theorem 4.1.1, we use Proposition 4.1.1 and Lemma 5.1.1 to obtain

$$\begin{aligned}
(\text{TI}) &\leq \sup_{\beta_0 \in \mathfrak{B}_{0,D}, \beta_1 \in \mathfrak{B}_1} \sup_{\beta_0' \in B(\beta_0)} \mathbb{P}_{0,\beta_0}\big(S(\mathcal{I}; \beta_0', \beta_1) \geq \tilde{t}_{\alpha-\gamma-\epsilon}(\beta_0', \beta_1)\big) + \delta \\
&\leq \sup_{\beta_0 \in \mathfrak{B}_{0,D}, \beta_1 \in \mathfrak{B}_1} \sup_{\beta_0' \in B(\beta_0)} \mathbb{P}_{0,\beta_0}\big(S(\mathcal{I}; \beta_0', \beta_1) \geq t_{\alpha-\gamma}(\beta_0', \beta_1)\big) \\
&\quad + \mathbb{P}_{0,\beta_0}\big(\tilde{t}_{\alpha-\gamma-\epsilon}(\beta_0', \beta_1) < t_{\alpha-\gamma}(\beta_0', \beta_1)\big) + \delta \\
&\leq \alpha - \gamma + \delta + \sup_{\beta_0 \in \mathfrak{B}_{0,D}, \beta_1 \in \mathfrak{B}_1} \sup_{\beta_0' \in B(\beta_0)} \mathbb{P}_{0,\beta_0}\big(\tilde{t}_{\alpha-\gamma-\epsilon}(\beta_0', \beta_1) < t_{\alpha-\gamma}(\beta_0', \beta_1)\big).
\end{aligned}$$

The second additional step that we need to take is to get this final probability back to a finite set of β_0 and β_1. By Lemma 5.3.2, we obtain

$$(\text{TI}) \leq \alpha - \gamma + \delta + \sup_{\beta_0 \in \mathfrak{B}_{0,D}, \beta_1 \in \mathfrak{B}_{1,D}} \mathbb{P}_{0,\beta_0}\Big(\hat{t}_{\alpha-\gamma-\epsilon}(\beta_0, \beta_1) < t_{\alpha-\gamma}(\beta_0, \beta_1)\Big).$$

Finally, by Proposition 5.1.2, we have

$$\begin{aligned}
(\text{TI}) &\leq \alpha - \gamma + \delta + \sum_{\beta_0 \in \mathfrak{B}_{0,D}, \beta_1 \in \mathfrak{B}_{1,D}} \mathbb{P}_{0,\beta_0}\Big(\hat{t}_{\alpha-\gamma-\epsilon}(\beta_0, \beta_1) < t_{\alpha-\gamma}(\beta_0, \beta_1)\Big) \\
&\leq \alpha - \gamma + \delta + \sum_{\beta_0 \in \mathfrak{B}_{0,D}, \beta_1 \in \mathfrak{B}_{1,D}} \frac{\xi}{N} \\
&= \alpha - \gamma + \delta + \xi \\
&\leq \alpha,
\end{aligned}$$

and this completes the proof. □

5.3.2 Two Statistic Test

The purpose of this section is to prove Corollary 4.2.1. The proof is similar to that of Theorem 4.1.1, although we need to be slightly more careful with the discretization since we do not require one of $-S_0(\mathcal{I})$ or $S_1(\mathcal{I})$ to be larger than all of their $1-(\alpha_i - \gamma - \epsilon)$ quantiles.

Proof of Proposition 4.2.1. Let $F_D : \mathfrak{B}_0 \to \mathfrak{B}_{0,D}$ be the discretization function. As in the Lipschitz case, we need a decoupling. Define the function

$$\hat{s}_{1,\alpha}(\beta) := \hat{s}_{1,\alpha,F_D(\beta')}.$$

Now, we split the desired probability using a union bound:

$$\begin{aligned}
(\text{TI}) &= \sup_{\beta_0 \in \mathfrak{B}_0} \mathbb{P}_{0,\beta_0}\Big(\{S_1(\mathcal{I}) > \hat{s}_{1,\alpha_1-\gamma-\epsilon}(\beta_0)\} \cup \Big\{S_0 < \hat{s}_{0,1-(\alpha_0-\gamma-\epsilon)}(\beta_0)\Big\}\Big) \\
&\leq \sup_{\beta_0 \in \mathfrak{B}_0} \mathbb{P}_{0,\beta_0}(S_1(\mathcal{I}) > \hat{s}_{1,\alpha_1-\gamma-\epsilon}(\beta_0)) + \sup_{\beta_0 \in \mathfrak{B}_0} \mathbb{P}_{0,\beta_0}\Big(S_0(\mathcal{I}) < \hat{s}_{0,1-(\alpha_0-\gamma-\epsilon)}(\beta_0)\Big).
\end{aligned}$$

Since the analyses of the S_0 and S_1 terms is identical up to a sign, we focus on the term involving S_1, which we denote by $(\mathrm{TI})_1$.

At this point, we observe that

$$\hat{s}_{1,\alpha_1-\gamma-\epsilon}(\beta) = \hat{s}_{1,\alpha_1-\gamma-\epsilon}(\beta')$$

whenever $F_D(\beta) = F_D(\beta')$. Thus, we have

$$\begin{aligned}
(\mathrm{TI})_1 &= \sup_{\beta_0 \in \mathfrak{B}_0} \mathbb{P}_{0,\beta_0}(S_1(\mathcal{I}) > \hat{s}_{1,\alpha_1-\gamma-\epsilon}(\beta_0)) \\
&= \sup_{\beta_0 \in \mathfrak{B}_{0,D}} \sup_{\beta' \in F_D^{-1}(\beta_0)} \mathbb{P}_{0,\beta'}(S_1(\mathcal{I}) > \hat{s}_{1,\alpha_1-\gamma-\epsilon}(\beta_0)) \\
&= \sup_{\beta_0 \in \mathfrak{B}_{0,D}} \sup_{\beta' \in F_D^{-1}(\beta_0)} \mathbb{P}_{0,\beta'}(S_1(\mathcal{I}) > \hat{s}_{1,\alpha_1-\gamma-\epsilon,\beta_0}) \\
&\leq \sup_{\beta_0 \in \mathfrak{B}_{0,D}} \mathbb{P}_{0,\beta_0}(S_1(\mathcal{I}) > \hat{s}_{1,\alpha_1-\gamma-\epsilon,\beta_0}) + \delta,
\end{aligned}$$

where the inequality is due to Proposition 4.1.1.

Now, the proof follows that of Theorem 4.1.1. Using Lemma 5.1.1, equation (2.1), and Proposition 5.1.2, we obtain

$$\begin{aligned}
(\mathrm{TI})_1 &= \sup_{\beta_0 \in \mathfrak{B}_{0,D}} \mathbb{P}_{0,\beta_0}(S_1(\mathcal{I}) > s_{1,\alpha_1-\gamma,\beta_0}) + \mathbb{P}_{0,\beta_0}(\hat{s}_{1,\alpha_1-\gamma,\beta_0} < s_{1,\alpha_1-\gamma,\beta_0}) + \delta \\
&\leq \alpha_1 - \gamma + \delta + \sup_{\beta_0 \in \mathfrak{B}_{0,D}} \mathbb{P}_{0,\beta_0}(\hat{s}_{1,\alpha_1-\gamma,\beta_0} < s_{1,\alpha_1-\gamma,\beta_0}) \\
&\leq \alpha_1 - \gamma + \delta + \sum_{\beta_0 \in \mathfrak{B}_{0,D}} \mathbb{P}_{0,\beta_0}(\hat{s}_{1,\alpha_1-\gamma,\beta_0} < s_{1,\alpha_1-\gamma,\beta_0}) \\
&\leq \alpha_1 - \gamma + \delta + \sum_{\beta_0 \in \mathfrak{B}_{0,D}} \frac{\xi}{N} \\
&\leq \alpha_1.
\end{aligned}$$

Since the term for S_0 may be computed similarly, we see that $(\mathrm{TI}) \leq \alpha_1 + \alpha_0 = \alpha$, which completes the proof. □

5.3.3 Multidimensional Weights

In this section, we want to prove Corollary 4.2.1. To this end, we need to prove the validity of the three parts of the discretization.

Discretizing $\mathfrak{C}_R$

The first step in this process is to prove Lemma 4.2.1.

Lemma 5.3.3. *Suppose that for $i = 1, \dots, d$, either $\beta_i \le \beta'_i$ or $\beta'_i \le \beta_i$. Define $\overline{\boldsymbol{\beta}} = \max\{\boldsymbol{\beta}, \boldsymbol{\beta}'\}$ and $\underline{\boldsymbol{\beta}} = \min\{\boldsymbol{\beta}, \boldsymbol{\beta}'\}$. Then, we have the total variation bound*

$$TV(\mathbb{P}_{\boldsymbol{\beta}}, \mathbb{P}_{\boldsymbol{\beta}'}) \le \frac{d+1}{2} \sum_{j=1}^{d} |\beta_j - \beta'_j| \min\left\{ \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \frac{W_t^j(P_{1:t-1})}{n+1-t}, \ \frac{k+c}{\underline{\beta}_j} \right\}.$$

Proof. As in the proof of Proposition 5.1.1, we want to define $\mathcal{I}$ and $\mathcal{I}'$ to have distributions governed by $\boldsymbol{\beta}$ and $\boldsymbol{\beta}'$, and we want $\mathcal{I}$ and $\mathcal{I}'$ to have simple descriptions. We start with $\mathcal{I}$. Again, let C be the random variable for the censored vertices. Let $\mathcal{P} = (\mathcal{P}_1, \dots, \mathcal{P}_{k+c})$ be a path of length $k+c$ with parameter vector $\boldsymbol{\beta}$, where $\mathcal{P}$ is defined by the random variables $X_{1:k+c}$ and $Y^i_{1:k+c}$ as follows. First, let X_t indicate the type of edge over which the spread occurred, with a uniform spread considered to be an edge of type 0. Specifically, let

$$\mathbb{P}(X_t = 0 | \mathcal{P}_{1:t-1} = P_{1:t-1}) = \frac{n+1-t}{n+1-t+W_t(P_{1:t-1})^\intercal \boldsymbol{\beta}}$$
$$\mathbb{P}(X_t = i | \mathcal{P}_{1:t-1} = P_{1:t-1}) = \frac{W_t^i(P_{1:t-1})\beta_i}{n+1-t+W_t(P_{1:t-1})^\intercal \boldsymbol{\beta}}.$$

Then, we define $Y_{t,i}$ to be the vertex that would be selected if $X_t = i$, which means that

$$\mathbb{P}(Y_{t,0} = v | \mathcal{P}_{1:t-1} = P_{1:t-1}) = \begin{cases} \frac{1}{n+1-t} & v \notin P_{1:t-1} \\ 0 & v \in P_{1:t-1} \end{cases}$$
$$\mathbb{P}(Y_{t,i} = v | \mathcal{P}_{1:t-1} = P_{1:t-1}) = \frac{W_{t,v}(P_{1:t-1})}{W_t(P_{1:t-1})}.$$

Thus, we have $\mathcal{P}_t = Y_{t,X_t}$. Then, we can write

$$\mathcal{I} = f(C, \mathcal{P}) = g(C, X_{1:k+c}, Y_{1:k+c,0:d}).$$

Similarly, we can define $X'_{1:k+c}$ and $(Y^{0:d}_{1:k+c})'$ for $\mathcal{I}'$. Note that since the $Y'_{t,i}$ do not depend on β_i conditioned on the path up to time t, we can set $Y^i_t = Y'_{t,i}$ as long as $X_{1:t-1} = X'_{1:t-1}$. Finally, we maximally couple the X_t and X'_t.

Applying Lemma 5.5.1 and Lemma 5.5.2, we have

$$\begin{aligned} \mathrm{TV}(\mathbb{P}_{\boldsymbol{\beta}}, \mathbb{P}_{\boldsymbol{\beta}'}) &\le \mathrm{TV}\Big(\mathbb{P}_{C, X_{1:k+c}, Y_{1:k+c,0:d}}, \mathbb{P}_{C, X'_{1:k+c}, Y'_{1:k+c,0:d}}\Big) \\ &\le \mathbb{P}\Big((C, X_{1:k+c}, Y_{1:k+c,0:d}) \neq (C, X'_{1:k+c}, Y'_{1:k+c,0:d})\Big) \\ &= \mathbb{P}(X_{1:k+c} \neq X_{1:k+c}). \end{aligned}$$

Now, we define the set $\mathfrak{P}_t$ to be

$$\mathfrak{P}_t := \Big\{ (X_{1:t}, Y_{1:t,0:d}, X'_{1:t}, Y'_{1:t,0:d}) : X_{1:t} = X'_{1:t} \Big\}.$$

Then, we have

$$\begin{aligned}\mathrm{TV}(\mathbb{P}_{\boldsymbol{\beta}}, \mathbb{P}_{\boldsymbol{\beta}'}) &\le \sum_{t=1}^{k+c} \mathbb{P}(X_{1:t-1} = X'_{1:t-1})\mathbb{P}(X_t \neq X'_t | X_{1:t-1} = X'_{1:t-1}) \\ &\le \sum_{t=1}^{k+c} \sum_{E \in \mathfrak{P}_{t-1}} \mathbb{P}(E)\mathbb{P}(X_t \neq X'_t | E) \\ &\le \sum_{t=1}^{k+c} \max_{E \in \mathfrak{P}_{t-1}} \mathbb{P}(X_t \neq X'_t | E).\end{aligned}$$

Let S_t be the tth summand in the above equation. Let $p_{t,i}$ and $q_{t,i}$ be the probability that X_t and X'_t are equal to i. Then, we can compute each summand as

$$S_t = \frac{1}{2} \max_{E \in \mathfrak{P}_{t-1}} \sum_{i=0}^{d} |p_{t,i} - q_{t,i}| \le \frac{1}{2} \sum_{i=0}^{d} \max_{E \in \mathfrak{P}_{t-1}} |p_{t,i} - q_{t,i}| =: \frac{1}{2} \sum_{i=0}^{d} \Delta_{t,i}.$$

Now, we can examine each of the $\Delta_{t,i}$. Consider $i > 0$. Let $\boldsymbol{\beta}^{(j)} = (\beta'_1, \ldots, \beta'_j, \beta_{j+1}, \ldots, \beta_d)$. Without loss of generality, consider $i = 1$, so that the ith index is the first to be changed in the telescoping that follows. We have

$$\begin{aligned}\Delta_{t,i} &= \max_{P_{1:t-1}} \left| \frac{W_t^i(P_{1:t-1})\beta_i}{n+1-t+W_t(P_{1:t-1})^\intercal \boldsymbol{\beta}} - \frac{W_t^i(P_{1:t-1})\beta'_i}{n+1-t+W_t(P_{1:t-1})^\intercal \boldsymbol{\beta}'} \right| \\ &= \max_{P_{1:t-1}} \left| \sum_{j=0}^{d} \frac{W_t^i(P_{1:t-1})\beta_i^{(j-1)}}{n+1-t+W_t(P_{1:t-1})^\intercal \boldsymbol{\beta}^{(j-1)}} - \frac{W_t^i(P_{1:t-1})\beta_i^{(j)}}{n+1-t+W_t(P_{1:t-1})^\intercal \boldsymbol{\beta}^{(j)}} \right| \\ &\le \sum_{j=0}^{d} \max_{P_{1:t-1}} \left| \frac{W_t^i(P_{1:t-1})\beta_i^{(j-1)}}{n+1-t+W_t(P_{1:t-1})^\intercal \boldsymbol{\beta}^{(j-1)}} - \frac{W_t^i(P_{1:t-1})\beta_i^{(j)}}{n+1-t+W_t(P_{1:t-1})^\intercal \boldsymbol{\beta}^{(j)}} \right| \\ &=: \sum_{j=1}^{d} \max_{P_{1:t-1}} \Delta_{t,i,j}.\end{aligned}$$

Thus, we need to bound each $\Delta_{t,i,j}$.

In the case that $i = j$, we have $\Delta_{t,i,i} = |h_i(\beta_i) - h_i(\beta'_i)|$ for h_i of the form $h_i(x) = ax/(b+ax)$. The derivative is

$$h'_i(x) = \left(\frac{a}{b+ax}\right)\left(\frac{b}{b+ax}\right) \le \min\left\{\frac{a}{b}, \frac{1}{2x}, \frac{b}{ax^2}\right\}.$$

Let the superscript $-i$ refer to a vector without its ith coordinate, e.g. for our parameter vector $\boldsymbol{\beta}$ in $\mathbb{R}^d$ we have $\boldsymbol{\beta}^{-i} = (\beta_1, \ldots, \beta_{i-1}, \beta_{i+1}, \ldots, \beta_d)$ in $\mathbb{R}^{d-1}$. Thus, we have

$$\begin{aligned}\Delta_{t,i,j} &\le \int_{\underline{\beta}_i}^{\overline{\beta}_i} \min\left\{\frac{W_t^i(P_{1:t-1})}{n+1-t+W_t^{-i}(P_{1:t-1})\boldsymbol{\beta}^{-i}}, \frac{1}{2x}, \frac{n+1-t+W^{-i}(P_{1:t-1})^\intercal \boldsymbol{\beta}^{-i}}{W_t^i(P_{1:t-1})x^2}\right\} dx \\ &\le \min\left\{|\beta_i - \beta'_i| \frac{W_t^i(P_{1:t-1})}{n+1-t}, \frac{1}{2}\log\frac{\overline{\beta}_i}{\underline{\beta}_i}, \frac{n+1-t+W_t^{-i}(P_{1:t-1})^\intercal \boldsymbol{\beta}^{-i}}{\underline{\beta}_i}\right\}.\end{aligned}$$

Consider the case of $i \neq j$. Then, note that we can write $\Delta_{t,i,j} = |h_{i,j}(\beta_j) - h_{i,j}(\beta'_j)|$ where $h_{i,j}$ has the form $h_{i,j}(x) = a/(b+cx)$. The derivative of this function satisfies the bound

$$|h'_{i,j}(x)| \leq \frac{ac}{(b+cx)^2} \leq \min\left\{\frac{ac}{b^2},\ \frac{a}{2bx},\ \frac{1}{cx^2}\right\}.$$

Thus by integrating, we have

$$\begin{aligned}
\Delta_{t,i,j} &\leq \int_{\underline{\beta}_j}^{\overline{\beta}_j} \min\left\{\frac{W_t^i(P_{1:t-1})\beta_i^{(j)}W_t^j(P_{1:t-1})}{(n+1-t+W_t^{-j}(P_{1:t-1})^\intercal\boldsymbol{\beta}^{-j})^2},\ \frac{W_t^i(P_{1:t-1})\beta_i^{(j)}}{2(n+1-t+W_t^{-j}(P_{1:t-1})^\intercal\boldsymbol{\beta}^{-j})x}\right\}dx \\
&\leq \int_{\underline{\beta}_j}^{\overline{\beta}_j} \min\left\{\frac{W_t^j(P_{1:t-1})}{n+1-t},\ \frac{1}{2x},\ \right\}dx \\
&\leq \min\left\{|\beta_j - \beta'_j|\frac{W_t^j(P_{1:t-1})}{n+1-t},\ \frac{1}{2}\log\frac{\overline{\beta}_j}{\underline{\beta}_j}\right\}.
\end{aligned}$$

So, we obtain

$$\Delta_{t,i} \leq \sum_{j=1}^{d} \max_{P_{1:t-1}} \min\left\{|\beta_j - \beta'_j|\frac{W_t^j(P_{1:t-1})}{n+1-t},\ \frac{1}{2}\log\frac{\overline{\beta}_j}{\underline{\beta}_j}\right\}.$$

Now, we consider $\Delta_{t,0}$. We can write

$$\Delta_{t,0} = |h_0(W_t(P_{1:t-1})^\intercal\boldsymbol{\beta}) - h_0(W_t(P_{1:t-1})^\intercal\boldsymbol{\beta}')|$$

where h_0 has the form $h_0(x) = a/(a+x)$. The derivative in this case satisfies the bound

$$|h'_0(x)| \leq \frac{a}{(a+x)^2}.$$

Integrating, we see that

$$\begin{aligned}
\Delta_{t,0} &\leq \max_{P_{1:t-1}} \int_{W_t(P_{1:t-1})^\intercal\underline{\boldsymbol{\beta}}}^{W_t(P_{1:t-1})^\intercal\overline{\boldsymbol{\beta}}} \frac{1}{n+1-t+x}dx \\
&\leq \max_{P_{1:t-1}} \frac{W_t(P_{1:t-1})^\intercal(\overline{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}})}{n+1-t+W_t(P_{1:t-1})^\intercal\underline{\boldsymbol{\beta}}} \\
&\leq \max_{P_{1:t-1}} \sum_{j=1}^{d} \min\left\{|\beta_j - \beta'_j|\frac{W_t^j(P_{1:t-1})}{n+1-t},\ \frac{\overline{\beta}_j - \underline{\beta}_j}{\underline{\beta}_j}.\right\}
\end{aligned}$$

Finally, putting everything together, we have

$$\begin{aligned}
\mathrm{TV}(\mathbb{P}_{\boldsymbol{\beta}}, \mathbb{P}_{\boldsymbol{\beta}'}) &\leq \sum_{t=1}^{k+c}\frac{1}{2}\sum_{i=0}^{d}\Delta_{t,i} \\
&\leq \frac{1}{2}\sum_{t=1}^{k+c}\sum_{i=0}^{d}\sum_{j=1}^{d}\max_{P_{1:t-1}}\min\left\{|\beta_j - \beta'_j|\frac{W_t^j(P_{1:t-1})}{n+1-t},\ \frac{\overline{\beta}_j - \underline{\beta}_j}{\underline{\beta}_j}.\right\} \\
&\leq \frac{d+1}{2}\sum_{j=1}^{d}|\beta_j - \beta'_j|\min\left\{\sum_{t=1}^{k+c}\max_{P_{1:t-1}}\frac{W_t^j(P_{1:t-1})}{n+1-t},\ \frac{k+c}{\underline{\beta}_j}\right\},
\end{aligned}$$

and this completes the proof of the lemma. □

Proof of Lemma 4.2.1. Let $\boldsymbol{\beta}' = F_{R,D}(\boldsymbol{\beta})$ for a $\boldsymbol{\beta}$ in $\mathfrak{C}_R$. By Lemma 5.3.3, we have

$$\mathrm{TV}(\mathbb{P}_{\boldsymbol{\beta}'}, \mathbb{P}_{\boldsymbol{\beta}}) \leq \frac{d+1}{2} \sum_{j=1}^{d} |\beta_j - \beta'_j| \min\left\{ \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \frac{W_t^j(P_{1:t-1})}{n+1-t},\ \frac{k+c}{\underline{\beta}_j} \right\}.$$

Now, for each i, we have

$$|\beta_i - \beta'_i| \min\left\{ \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \frac{W_t^i(P_{1:t-1})}{n+1-t},\ \frac{k+c}{\underline{\beta}_i} \right\} \leq \frac{2\delta}{d(d+1)},$$

where the first term in the minimum is used for large β_i and the second term is used for small β_i. Thus, we have

$$\mathrm{TV}(\mathbb{P}_{\boldsymbol{\beta}'}, \mathbb{P}_{\boldsymbol{\beta}}) \leq \frac{d+1}{2} \sum_{i=1}^{d} \frac{2\delta}{d(d+1)} = \delta.$$

proving the lemma. □

Discretizing with Δ_D^2

Now, we have two more parts of the discretization to prove. Next, we consider the case of $\boldsymbol{\eta}$ in Δ^2 by proving Lemma 4.2.2.

Proof of Lemma 4.2.2. For $\boldsymbol{\beta} = (\beta_1, \beta_2)$ where both entries are large, we shall show the probability distribution is sufficiently close to the distribution given by a $\boldsymbol{\eta}$ in Δ^2, which is close enough to the distribution parametrized by a $\boldsymbol{\eta}'$ in Δ_D^2. Thus, we start with the discretization of Δ^2 for the second approximation step.

First, we wish to show that for every $\boldsymbol{\eta}$ in Δ^2, there is an $\boldsymbol{\eta}'$ in Δ_D^2 such that

$$\mathrm{TV}(\mathbb{P}_{\boldsymbol{\eta}}, \mathbb{P}_{\boldsymbol{\eta}'}) \leq \frac{\delta}{2}. \tag{5.2}$$

We shall prove this by a coupling argument.

Let $\mathcal{I}$ and $\mathcal{I}'$ be infections according to $\boldsymbol{\eta}$ and $\boldsymbol{\eta}'$ respectively. They shall be coupled in the following manner. First, choose a set C of c vertices to be censored by an arbitrary mechanism, as long as it is independent of the ensuing path variables. Next, define the variable X_t to be an indicator random variable that is 1 if

(a) at time t, there are no edges between infected and uninfected vertices,

(b) at time t, there are vertices of only a single type, e.g. $W_t^1(P_{1:t-1}) = 0$ or $W_t^2(P_{1:t-1}) = 0$, or

(c) at time t, the spread occurs over an edge of type i where i is the minimizer in the definition of M_∞.

For case (c), assume without loss of generality that $i = 1$, and note that this occurs with probability

$$\mathbb{P}(X_t = 1 | \mathcal{P}_{1:t-1} = P_{1:t-1}) = \frac{W_t^1(P_{1:t-1})\eta_1}{W_t(P_{1:t-1}) + (W_t^1(P_{1:t-1}) - W_t^2(P_{1:t-1}))\eta_1}$$

The only case where $X_t = 0$ is when there are edges of both types, i.e. $W_t^1, W_t^2 > 0$, and yet a spread over an edge of type two occurs. Let A_t be the infected vertex in case (a), let B_t be the infected vertex in case (b), and let C_t be the infected vertex in case (c). Let D_t be the infected vertex when $X_t = 0$. Note that A_t, B_t, C_t, and D_t can be chosen independently of X_t, and then the vertex that ultimately becomes infected is added to the path as $\mathcal{P}_t$. Finally, from C and $\mathcal{P}$ we can define the infection vector $\mathcal{I}$ as

$$\mathcal{I} = f(C, \mathcal{P}) = g(C, X_1, A_{1:k+c}, B_{1:k+c}, C_{1:k+c}, D_{1:k+c})$$

for some functions f and g.

At this point, we want to define $\mathcal{I}'$. So, we let X_t', A_t', B_t', C_t', D_t', be defined analogously. Then, we can write

$$\mathcal{I}' = f(C, \mathcal{P}') = g(C, X'_{1:k+c}, A'_{1:k+c}, B'_{1:k+c}, C'_{1:k+c}, D'_{1:k+c}).$$

Further, we want the following coupling. In the case that $\mathcal{P}_{1:t-1} = \mathcal{P}'_{1:t-1}$, we can set $(A_t, B_t, C_t) = (A_t', B_t', C_t')$. Additionally, we maximally couple X_t and X_t'.

Now, by Lemma 5.5.1 and Lemma 5.5.2, we have

$$\begin{aligned}
\mathrm{TV}(\mathbb{P}_{\boldsymbol{\eta}}, \mathbb{P}_{\boldsymbol{\eta}'}) &\le \mathrm{TV}\Big(\mathbb{P}_{C, X_{1:k+c}, A_{1:k+c}, B_{1:k+c}, C_{1:k+c}, D_{1:k+c}}, \mathbb{P}_{C, X'_{1:k+c}, A'_{1:k+c}, B'_{1:k+c}, C'_{1:k+c}, D'_{1:k+c}}\Big) \\
&\le \mathbb{P}((C, X_{1:k+c}, A_{1:k+c}, B_{1:k+c}, C_{1:k+c}, D_{1:k+c}) \\
&\qquad \neq (C, X'_{1:k+c}, A'_{1:k+c}, B'_{1:k+c}, C'_{1:k+c}, D'_{1:k+c})) \\
&= \mathbb{P}((X_1, \dots, X_{k+c}) \neq (X'_1, \dots, X'_{k+c})).
\end{aligned}$$

Let $\mathfrak{P}_t$ denote the set

$$\mathfrak{P}_t = \{(X_{1:t}, A_{1:t}, B_{1:t}, C_{1:t}, D_{1:t}, X'_{1:t}, A'_{1:t}, B'_{1:t}, C'_{1:t}, D'_{1:t}) : X_{1:t} = X'_{1:t}\}.$$

Then, we have the upper bound

$$\begin{aligned}
\mathrm{TV}(\mathbb{P}_{\boldsymbol{\eta}}, \mathbb{P}_{\boldsymbol{\eta}'}) &\le \sum_{t=1}^{k+c} \sum_{E \in \mathfrak{P}_{t-1}} \mathbb{P}(E)\mathbb{P}(X_t \neq X_t' | E) \\
&\le \sum_{t=1}^{k+c} \max_{E \in \mathfrak{P}_{t-1}} \mathbb{P}(X_t \neq X_t' | E) \\
&\le \sum_{t=1}^{k+c} \max_{E \in \mathfrak{P}_{t-1}} \Big| \frac{W_t^1(P_{1:t-1})\eta_1}{W_t^2(P_{1:t-1}) + [W_t^1(P_{1:t-1}) - W_t^2(P_{1:t-1})]\eta_1} \\
&\qquad - \frac{W_t^1(P_{1:t-1})\eta_1'}{W_t^2(P_{1:t-1}) + [W_t^1(P_{1:t-1}) - W_t^2(P_{1:t-1})]\eta_1'} \Big|.
\end{aligned}$$

Now, we need to bound the differences in the summands. To this end, we shall use a Lipschitz result from examining the derivative. Define the function

$$g(x) = \frac{ax}{b + (a-b)x},$$

where $a \geq 1$ and $b \geq 1$ are positive constants. Then, the derivative of g is

$$g'(x) = \frac{a}{b + (a-b)x} \cdot \frac{b}{b + (a-b)x},$$

and so g is a-Lipschitz. Applying this to our present problem, we see that

$$\mathrm{TV}(\mathbb{P}_{\boldsymbol{\eta}}, \mathbb{P}_{\boldsymbol{\eta}'}) \leq |\eta_1 - \eta_1'| \sum_{t=1}^{k+c} \max_{P_{1:t-1}} W_t^1(P_{1:t-1}) \leq \frac{\delta}{2},$$

which proves the validity of the discretization for $\boldsymbol{\eta}$ in Δ^2.

Now, we consider the case where $\beta_1, \beta_2 \geq K$. We need to make another coupling argument. Again let C be the censored vertices. At this point, we redefine X_t to be the indicator that

(a) there are no edges over which to spread or

(b) spreading occurs over an edge of either type one or two.

Should case (a) occur, define the resulting infected vertex to be A_t. Should case (b) occur, define the resulting infected vertex to be B_t. If $X_t = 0$, then spreading did not occur over an edge even though it was possible, and define the resulting infected vertex to be C_t). As before, we can write

$$\mathcal{I} = f(C, \mathcal{P}) = g(C, X_{1:k+c}, A_{1:k+c}, B_{1:k+c}, C_{1:k+c}).$$

Again, we define X_t' to be the analogous quanitity for $\mathcal{I}'$, and we can couple the variables so that $(A_t', B_t', C_t') = (A_t, B_t, C_t)$ when $X_{1:t-1} = X_{1:t-1}'$. Note that in this case, $X_t' = 1$ almost surely. Similarly, we have

$$\mathcal{I}' = f(C, \mathcal{P}') = g(C, X_{1:k+c}', A_{1:k+c}', B_{1:k+c}', C_{1:k+c}').$$

Again using Lemma 5.5.1 and Lemma 5.5.2, we have

$$\begin{aligned}
\mathrm{TV}(\mathbb{P}_{\boldsymbol{\beta}}, \mathbb{P}_{\boldsymbol{\eta}}) &\leq \sum_{t=1}^{k+c} \max_{E \in \mathfrak{P}_{t-1}} \mathbb{P}(X_t \neq X_t' | E) \\
&= \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \frac{n+1-t}{n+1-t+W_t(P_{1:t-1})^\intercal \boldsymbol{\beta}} \mathbf{1}\Big\{W_t^1(P_{1:t-1}) + W_t^2(P_{1:t-1}) \geq 1\Big\}.
\end{aligned}$$

Using the fact that $\beta_1, \beta_2 \geq K$ and we assume the minimal non-zero edge weights to be at least 1, we have

$$\mathrm{TV}(\mathbb{P}_{\boldsymbol{\beta}}, \mathbb{P}_{\boldsymbol{\eta}}) \leq \sum_{t=1}^{k+c'} \frac{n+1-t}{n+1-t+K} \leq \frac{\delta}{2}.$$

Using our earlier result on the discretization of Δ^2, we have

$$\mathrm{TV}(\mathbb{P}_{\boldsymbol{\beta}}, \mathbb{P}_{\boldsymbol{\eta}'}) \leq \mathrm{TV}(\mathbb{P}_{\boldsymbol{\beta}}, \mathbb{P}_{\boldsymbol{\eta}}) + \mathrm{TV}(\mathbb{P}_{\boldsymbol{\eta}}, \mathbb{P}_{\boldsymbol{\eta}'}) \leq \delta,$$

which is what we want. □

Discretizing with $\mathfrak{H}_D$

Finally, we consider our half-infinite $\boldsymbol{\beta}$ discretization. To prove Lemma 4.2.3, we first provide a more general lemma.

Lemma 5.3.4. *Let $\mathbb{P}_{\boldsymbol{\beta}}$ and $\mathbb{P}_{\zeta}$ be the measures of $\mathcal{I}$ and $\mathcal{S}$. Then, we have*

$$TV(\mathbb{P}_{\zeta}, \mathbb{P}_{\boldsymbol{\beta}}) \leq \frac{N_0 + M_0^1\beta_1}{\beta_2} + \min\left\{|\beta_1 - \zeta| M_0^1,\ \frac{k+c}{2}\log\frac{\max\{\beta_1,\zeta\}}{\min\{\beta_1,\zeta\}},\ \frac{N_0}{\min\{\beta_1,\zeta\}}\right\}.$$

Proof. Without loss of generality, assume that $\beta_1 \leq \beta_2$. We describe our construction of $\mathcal{I}$ and $\mathcal{S}$ so that they are spreads over $\boldsymbol{\beta}$ and ζ respectively. We denote the infection paths they form by $\mathcal{P}$ and $\mathcal{P}'$ respectively. First, let C denote the c censored vertices, which may be chosen arbitrarily for our purposes as long as the mechanism of the choice is the same for $\mathcal{I}$ and $\mathcal{S}$. Define $X_{t,2}$ to be the indicator of a spread over an edge of type two at time t, i.e.

$$\mathbb{P}(X_{t,2} = 1 | \mathcal{P}_{1:t-1} = P_{1:t-1}) = \frac{W_t^2(P_{1:t-1})\beta_2}{n+1-t+W_t^1(P_{1:t-1})\beta_1 + W_t^2(P_{1:t-1})\beta_2}.$$

Define $X_{t,1}$ to be the indicator of a apread over an edge of type one at time t conditioned on no spread over an edge of type two, i.e.

$$\mathbb{P}(X_{t,1} = 1 | \mathcal{P}_{1:t-1} = P_{1:t-1}) = \frac{W_t^1(P_{1:t-1})\beta_1}{n+1-t+W_t^1(P_{1:t-1})\beta_1}.$$

Define $Y_{t,i}$ for $i = 0, 1, 2$ to be the infected vertex if spread occurs over an edge of type i, where an edge of type 0 is considered to be uniform spread. Specifically, we have

$$\mathbb{P}(Y_{t,0} = v | \mathcal{P}_{1:t-1} = P_{1:t-1}) = \begin{cases} \frac{1}{n+1-t} & v \notin P_{1:t-1} \\ 0 & v \in P_{1:t-1} \end{cases}$$

$$\mathbb{P}(Y_{t,i} = v | \mathcal{P}_{1:t-1} = P_{1:t-1}) = \frac{W_{t,v}^i(P_{1:t-1})}{W_t^i(P_{1:t-1})} \quad i = 1, 2.$$

From these random variables, we define $\mathcal{P}_t = Y_{t,2}$ when $X_{t,2} = 1$, $\mathcal{P}_t = Y_{t,1}$ if $X_{t,2} = 0$ and $X_{t,1} = 1$, and $\mathcal{P}_t = Y_{t,0}$ if $X_{t,2} = X_{t,1} = 0$. Thus, we can write

$$\mathcal{I} = f(C, \mathcal{P}) = g(C, X_{1:k+c,1:2}, Y_{1:k+c,0:2}).$$

Here, we can think of the function f in the following manner: the infected vertices k infected vertices are the first k infected vertices of $\mathcal{P}$ that are not censored, i.e. in C. For g, note that by choosing the $X_{t,i}$ and $Y_{t,i}$, we can reconstruct $\mathcal{P}$.

We analogously define $X'_{t,i}$ and $Y'_{t,i}$ for the process $\mathcal{S}$. If $\mathcal{P}_{1:t-1} = \mathcal{P}'_{1:t-1}$, then set $Y'_{t,i} = Y_{t,i}$ for each i. Finally, note that we can set

$$\mathcal{S} = g(C, X'_{1:k+c,1:2}, Y'_{1:k+c,0:2}).$$

The next step is to use the data processing inequality of Lemma 5.5.1 and Lemma 5.5.2 to obtain

$$\begin{aligned}
\mathrm{TV}(\mathbb{P}_\zeta, \mathbb{P}_{\boldsymbol{\beta}}) &\leq \mathrm{TV}\Big(\mathbb{P}_{C,X_{1:k+c,1:2},Y_{1:k+c,0:2}}, \mathbb{P}_{C,X'_{1:k+c,1:2},Y'_{1:k+c,0:2}}\Big) \\
&\leq \mathbb{P}\Big((C, X_{1:k+c,1:2}, Y_{1:k+c,0:2}) \neq (C, X'_{1:k+c,1:2}, Y'_{1:k+c,0:2})\Big) \\
&= \mathbb{P}\Big(X_{1:k+c,1:2} \neq X'_{1:k+c,1:2}\Big).
\end{aligned}$$

Now, we define the set $\mathfrak{P}_t$ to be

$$\mathfrak{P}_t := \Big\{(X_{1:t,1:2}, Y_{1:t,0:2}, X'_{1:t,1:2}, Y'_{1:t,0:2}), : X_{1:t,1:2} = X'_{1:t,1:2}\Big\}.$$

Then, we have

$$\begin{aligned}
\mathrm{TV}(\mathbb{P}_\zeta, \mathbb{P}_{\boldsymbol{\beta}}) &\leq \sum_{t=1}^{k+c} \mathbb{P}\Big(X_{1:t-1,1:2} = X'_{1:t-1,1:2}\Big)\mathbb{P}\Big(X_{t,1:2} \neq X'_{t,1:2} | X_{1:t-1,1:2} = X'_{1:t-1,1:2}\Big) \\
&\leq \sum_{t=1}^{k+c} \sum_{E \in \mathfrak{P}_{t-1}} \mathbb{P}(E)\mathbb{P}\Big(X_{t,1:2} \neq X'_{t,1:2} | E\Big) \\
&\leq \sum_{t=1}^{k+c} \max_{E \in \mathfrak{P}_{t-1}} \mathbb{P}\Big(X_{t,1:2} \neq X'_{t,1:2} | E\Big) \\
&\leq \sum_{t=1}^{k+c} \max_{E \in \mathfrak{P}_{t-1}} \mathbb{P}\Big(X_{t,1} \neq X'_{t,1} | E\Big) + \sum_{t=1}^{k+c} \max_{E \in \mathfrak{P}_{t-1}} \mathbb{P}\Big(X_{t,2} \neq X'_{t,2} | E\Big) \\
&=: S_1 + S_2.
\end{aligned}$$

So, it suffices to bound these sums S_1 and S_2.

We start with the S_1. We have

$$\begin{aligned}
S_1 &= \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \left| \frac{W_t^2(P_{1:t-1})\beta_2}{n+1-t+W_t^1(P_{1:t-1})\beta_1 + W_t^2(P_{1:t-1})\beta_2} - 1 \right| \mathbf{1}\Big\{W_t^2(P_{1:t-1}) > 0\Big\} \\
&\leq \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \frac{n+1-t+W_t^1(P_{1:t-1})\beta_1}{n+1-t+W_t^1(P_{1:t-1})\beta_1 + W_t^2(P_{1:t-1})\beta_2} \mathbf{1}\Big\{W_t^2(P_{1:t-1}) > 0\Big\} \\
&\leq \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \frac{n+1-t+W_t^1(P_{1:t-1})\beta_1}{\beta_2} \\
&= \frac{1}{\beta_2}\left((k+c)\Big(n - \frac{k+c-1}{2}\Big) + \beta_1 \sum_{t=1}^{k+c} \max_{P_{1:t-1}} W_t^1(P_{1:t-1})\right) \\
&= \frac{N_0 + M_0^1\beta_1}{\beta_2}.
\end{aligned}$$

Next, we consider the bound for S_2. We have

$$\begin{aligned} S_2 &= \sum_{t=1}^{k+c} \max_{P_{1:t-1}} \left| \frac{W_t^1(P_{1:t-1})\beta_1}{n+1-t+W_t^1(P_{1:t-1})\beta_1} - \frac{W_t^1(P_{1:t-1})\zeta}{n+1-t+W_t^1(P_{1:t-1})\zeta} \right| \\ &\leq \min\left\{ |\beta_1 - \zeta| M_0,\ \frac{k+c}{2} \log \frac{\max\{\beta_1, \zeta\}}{\min\{\beta_1, \zeta\}},\ \frac{N_0}{\min\{\beta_1, \zeta\}} \right\}, \end{aligned}$$

where the inequality comes from analyzing the same quantity in Proposition 5.1.1. Thus, putting everything together completes the proof. □

With this lemma in hand, we can prove Lemma 4.2.3.

Proof of Lemma 4.2.3. Consider $\boldsymbol{\beta} = (\beta_1, \beta_2)$ with $\beta_1 \leq \beta_2$. Then, by assumption, we have

$$\frac{N_0 + M_0^1 \beta_1}{\beta_2} \leq \frac{N_0 + M_0^1 \beta_1}{\frac{2}{\delta}(N_0 + M_0^1 \beta_1)} = \frac{\delta}{2}.$$

By our discretization, as in the proof of Proposition 4.1.1, we have

$$\min\left\{ |\beta_1 - \zeta| M_0,\ \frac{k+c}{2} \log \frac{\max\{\beta_1, \zeta\}}{\min\{\beta_1, \zeta\}},\ \frac{N_0}{\min\{\beta_1, \zeta\}} \right\} \leq \frac{\delta}{2}.$$

Thus, by Lemma 5.3.4, we have

$$\mathrm{TV}(\mathbb{P}_{\boldsymbol{\beta}}, \mathbb{P}_{\zeta}) \leq \delta,$$

as desired. □

5.3.4 Confidence Sets

In this section, we give an additional confidence set algorithm. This algorithm mirrors the two statistic test of Algorithm 4.2.1.

Corollary 5.3.1. *The set C given by Algorithm 5.3.2 is a $(1-\alpha)$-confidence set.*

5.4 Ising Model

In this section, we consider a discretization for the Ising model. The discretization is only necessary when there is censoring, since for $c = 0$, we can apply the monotonicity argument of Chapter 3.

As usual, we start by defining a discretization. As in the case of multidimensional weights for $d > 2$, our discretization only applies for bounded parameter sets $\mathfrak{B} \subseteq \mathfrak{C}_{\mathrm{Ising},R} = [0, R]$.

Definition 5.4.1. *Let $H = \max_{J \in \mathcal{U}(\mathbb{I}_{k,c})} |\mathcal{H}(J)|$. Define*

$$N = \left\lceil \frac{RH}{\delta^2} \right\rceil + 1$$

Algorithm 5.3.2: Two-Statistic Confidence Set

Input : Type I error tolerances α_0 and α_1 where $\alpha_0 + \alpha_1 = \alpha$, approximation parameters ϵ, δ, γ, and ξ, observed infection vector $\mathcal{I}$, null graph $\mathcal{G}$, statistics S_0 and S_1, discretization $\mathfrak{B}_D$ of size N, discretization function F_D.

1 Define $N_{\text{sims}} \geq \left(\frac{1}{2\epsilon^2} + \frac{8}{3\epsilon}\right) \log \frac{N}{\xi}$.

2 For each β in $\mathfrak{B}_D$, simulate an infection N_{sims} times on $\mathcal{G}$ to obtain the approximate the $(\alpha_0 - \gamma - \epsilon)$ quantile $\hat{s}_{0,1-\alpha_0-\gamma-\epsilon,\beta}$ and the $1 - (\alpha_1 - \gamma - \epsilon)$ quantile $\hat{s}_{1,(\alpha_1-\gamma-\epsilon),\beta}$.

3 Define the discrete confidence set

$$C_D := \left\{\beta \in \mathfrak{B}_D : S_0(\mathcal{I}) \geq \hat{s}_{0,1-\alpha_0-\gamma-\epsilon,\beta} \text{ and } S_1(\mathcal{I}) \leq \hat{s}_{1,(\alpha_1-\gamma-\epsilon),\beta}\right\}.$$

4 Return the confidence set $C = F_D^{-1}(C_D)$.

Then, we define the discretization

$$\mathfrak{C}_{Ising,R,D} := \left\{\frac{(i-1)\delta^2}{H} : i = 1, \ldots, N\right\}.$$

The associated discretization function is

$$F_{Ising,D}(\beta) = \underset{\beta' \in \mathfrak{C}_{Ising,D}}{\arg\min} \ |\beta - \beta'|.$$

Proposition 5.4.1. *Let $\mathfrak{B}_{Ising}$ be a parameter set bounded by R. The discretization $\mathfrak{C}_{Ising,R,D}$ with discretization function $F_{Ising,D}$ satisfies equation (4.1).*

We shall now embark on proving this proposition. Note that because we lack a simple stochastic process representation here, we need to use different techniques than for the stochastic spreading model. First, we consider a helpful lemma.

Lemma 5.4.1. *Let $S : \mathbb{I}_{k,c} \to \mathbb{R}$ be a statistic. Consider the generalized Ising model with energy function $\mathcal{H}$. Let $H = \max_{J \in \mathcal{U}(\mathbb{I}_{k,c})} |\mathcal{H}(J)|$. Then, we have*

$$TV(\mathbb{P}_{Ising,\beta}, \mathbb{P}_{Ising,\beta'}) \leq \sqrt{2H|\beta - \beta'|}.$$

Proof. We start with using Pinsker's inequality to obtain

$$\mathrm{TV}(\mathbb{P}_{\mathrm{Ising},\beta}, \mathbb{P}_{\mathrm{Ising},\beta'}) \leq \sqrt{\frac{1}{2}\mathrm{KL}(\mathbb{P}_{\mathrm{Ising},\beta}, \mathbb{P}_{\mathrm{Ising},\beta'})}.$$

Thus, it only remains to compute the Kullback-Leibler divergence. Now, we simply need to compute and use the log-sum inequality of Lemma 5.5.4 repeatedly. We have

$$\begin{aligned}
\mathrm{KL}(\mathbb{P}_{\mathrm{Ising},\beta}, \mathbb{P}_{\mathrm{Ising},\beta'}) &= \sum_{J \in \mathbb{I}_{k,c}} \mathbb{P}_{\mathrm{Ising},\beta}(\mathcal{I} = J) \log \frac{\mathbb{P}_{\mathrm{Ising},\beta}(\mathcal{I} = J)}{\mathbb{P}_{\mathrm{Ising},\beta'}(\mathcal{I} = J)} \\
&= \sum_{J \in \mathbb{I}_{k,c}} \frac{1}{Z(\beta)} \sum_{J' \in \mathcal{U}(J)} \frac{1}{Z_{\mathrm{Ising}}(\beta)} \exp(-\beta \mathcal{H}(J')) \\
&\qquad \times \left(\log \frac{Z(\beta')}{Z(\beta)} + \log \frac{\sum_{J \in \mathcal{U}(J)} \frac{1}{Z_{\mathrm{Ising}}(\beta)} \exp(-\beta \mathcal{H}(J'))}{\sum_{J' \in \mathcal{U}(J)} \frac{1}{Z_{\mathrm{Ising}}(\beta')} \exp(-\beta' \mathcal{H}(J'))} \right) \\
&\leq \log \frac{Z(\beta')}{Z(\beta)} + \sum_{J \in \mathbb{I}_{k,c}} \sum_{J' \in \mathcal{U}(J)} \frac{1}{Z(\beta)} \cdot \frac{1}{Z_{\mathrm{Ising}}(\beta)} \exp(-\beta \mathcal{H}(J')) \\
&\qquad \times \left(\log \frac{Z_{\mathrm{Ising}}(\beta')}{Z_{\mathrm{Ising}}(\beta)} + \log \frac{\exp(-\beta \mathcal{H}(J'))}{\exp(-\beta' \mathcal{H}(J'))} \right) \\
&\leq \log \frac{Z(\beta')}{Z(\beta)} \\
&\qquad + \sum_{J \in \mathbb{I}_{k,c}} \sum_{J' \in \mathcal{U}(J)} \frac{1}{Z(\beta)} \cdot \frac{\exp(-\beta \mathcal{H}(J'))}{Z_{\mathrm{Ising},k+c'}(\beta)} \log \frac{Z_{\mathrm{Ising},k+c'}(\beta')}{Z_{\mathrm{Ising},k+c'}(\beta)} \\
&\qquad + \sum_{J \in \mathbb{I}_{k,c}} \sum_{J' \in \mathcal{U}(J)} \frac{1}{Z(\beta)} \cdot \frac{\exp(-\beta \mathcal{H}(J'))}{Z_{\mathrm{Ising}}(\beta)} \mathcal{H}(J')(\beta' - \beta)) \\
&=: S_1 + S_2 + (\beta' - \beta) \mathbb{E}_{\mathrm{Ising},\beta} \mathcal{H}(\mathcal{I})
\end{aligned}$$

Note that the final term can be bounded by $|\beta' - \beta| H$. Now, we need to analyze the first two terms on the right hand side above, which we denote S_1 and S_2. Starting with the latter, we have

$$\begin{aligned}
S_2 &= \sum_{J \in \mathbb{I}_{k,c}} \sum_{J' \in \mathcal{U}(J)} \frac{1}{Z(\beta)} \cdot \frac{1}{Z_{\mathrm{Ising},k+c'}(\beta)} \exp(-\beta \mathcal{H}(J')) \log \frac{\sum_{J \in \mathbb{I}_{k+c',0}} \exp(-\beta' \mathcal{H}(J))}{\sum_{J \in \mathbb{I}_{k+c',0}} \exp(-\beta \mathcal{H}(J))} \\
&\leq \sum_{J \in \mathbb{I}_{k,c}} \sum_{J' \in \mathcal{U}(J)} \frac{1}{Z(\beta)} \cdot \frac{1}{Z_{\mathrm{Ising},k+c'}(\beta)} \exp(-\beta \mathcal{H}(J')) \\
&\qquad \times \frac{1}{Z_{\mathrm{Ising}}(\beta')} \sum_{J \in \mathbb{I}_{k+c',0}} \exp(-\beta' \mathcal{H}(J)) \log \frac{\exp(-\beta' \mathcal{H}(J))}{\exp(-\beta \mathcal{H}(J))} \\
&= \sum_{J \in \mathbb{I}_{k,c}} \sum_{J' \in \mathcal{U}(J)} \frac{1}{Z(\beta)} \cdot \frac{1}{Z_{\mathrm{Ising},k+c'}(\beta)} \exp(-\beta \mathcal{H}(J')) \\
&\qquad \times \frac{1}{Z_{\mathrm{Ising}}} \sum_{J \in \mathbb{I}_{k+c',0}} \exp(-\beta' \mathcal{H}(J)) \mathcal{H}(J)(\beta - \beta') \\
&\leq \sum_{J \in \mathbb{I}_{k,c}} \sum_{J' \in \mathcal{U}(J)} \frac{1}{Z(\beta)} \cdot \frac{1}{Z_{\mathrm{Ising},k+c'}(\beta)} \exp(-\beta \mathcal{H}(J')) |\beta - \beta'| H \\
&= |\beta - \beta'| H.
\end{aligned}$$

Now, we can compute S_1. We have

$$\begin{aligned}
S_1 &= \log \frac{Z(\beta')}{Z(\beta)} \\
&\leq \frac{1}{Z(\beta')} \sum_{J \in \mathbb{I}_{k,c}} \sum_{J' \in \mathcal{U}(J)} \frac{1}{Z_{\text{Ising},k+c'}(\beta')} \exp(-\beta' \mathcal{H}(J')) \\
&\qquad \times \left(\log \frac{Z_{\text{Ising},k+c'}(\beta)}{Z_{\text{Ising},k+c'}(\beta')} + \log \frac{\exp(-\beta' \mathcal{H}(J'))}{\exp(-\beta \mathcal{H}(J'))} \right) \\
&\leq |\beta - \beta'| H + (\beta - \beta') \mathbb{E}_{\text{Ising},\beta'} \mathcal{H}(\mathcal{I}) \\
&\leq 2|\beta - \beta'| H.
\end{aligned}$$

Putting everything together completes the proof. □

Proof of Proposition 5.4.1. Let β in $\mathfrak{C}_{\text{Ising}}$ be given, and define $\beta' = F_{\text{Ising},D}(\beta)$. Thus, by our discretization, we have

$$|\beta - \beta'| \leq \frac{\delta^2}{2H}.$$

Plugging this into the result of Lemma 5.4.1 completes the proof. □

5.5 Auxiliary Lemmas

In this section, we provide a few standard lemmas that are useful in our analysis. First, we provide a data-processing inequality of Wu (2017). For this, let $D_f(P, Q)$ denote the f-divergence of P and Q, which is defined as

$$D_f(P, Q) := \mathbb{E}_D f\left(\frac{P}{Q}\right)$$

for an f satisfying specific properties. For our purposes, we are interested in $f(t) = t \log t$ and $f(t) = |t - 1|/2$, which correspond to the Kullback-Leibler divergence and the total variation distance respectively.

Lemma 5.5.1. *Let X be a random variable with distributions P_X and Q_X. Let $Y = g(X)$ for some function g, and define the induced distributions of Y to be P_X and Q_Y. Then, we have*

$$D_f(P_Y, Q_Y) \leq D_f(P_X, Q_X).$$

Proof. For the proof, we use p_x, p_y, and $p_{x|y}$ to refer to probability masses for $X = x$, $Y = y$, and $X = x$ given $Y = y$. Additionally, we define the analogous quantities for q.

Now, we proceed with the proof. Since x fully determines y, we have

$$\begin{aligned} D_f(P_X, Q_X) &= \sum_x q_x f\left(\frac{p_x}{q_x}\right) \\ &= \sum_{x,y} q_{x,y} f\left(\frac{p_{x,y}}{q_{x,y}}\right) \\ &= \sum_y q_y \sum_x q_{x|y} f\left(\frac{p_{xy}}{q_{xy}}\right) \\ &= \mathbb{E}_{Q_Y} \mathbb{E}_{Q_{X|Y}} f\left(\frac{P_{XY}}{Q_{XY}}\right). \end{aligned}$$

Presently, we can use Jensen's inequality, which yields

$$\begin{aligned} D_f(P_X, Q_X) &\geq \mathbb{E}_{Q_Y} f\left(\mathbb{E}_{Q_{X|Y}} \frac{P_{XY}}{Q_{XY}}\right) \\ &= \sum_y q_y f\left(\sum_x q_{x|y} \frac{p_{xy}}{q_{xy}}\right) \\ &= \sum_y q_y f\left(\sum_x \frac{p_y p_{x|y}}{p_y}\right) \\ &= \sum_y q_y f\left(\frac{p_y}{p_y} \sum_x p_{x|y}\right) \\ &= \sum_y q_y f\left(\frac{p_y}{p_y}\right) \\ &= D_f(P_Y, Q_Y). \end{aligned}$$

This completes the proof. □

Lemma 5.5.2. *Let X and X' be jointly-defined random variables on the same finite space $\mathcal{X}$ with respect to the measure $\mathbb{P}$. Let the respective marginal measures be $\mathbb{P}_\beta$ and $\mathbb{P}_{\beta'}$, i.e. $\mathbb{P}_\beta(X = J)$ and $\mathbb{P}_{\beta'}(X' = J)$ are the marginal distributions. Then, we have*

$$TV(\mathbb{P}_\beta, \mathbb{P}_{\beta'}) \leq \mathbb{P}(X \neq X').$$

Proof of Lemma 5.5.2. Let E be any event. Then, E has the form

$$E = \{\mathcal{P} \in E'\}$$

for some subset E' of the space of possible outcomes $\mathcal{X}$. By definition, we have

$$\mathrm{TV}(\mathbb{P}_\beta, \mathbb{P}_{\beta'}) = \sup_E |\mathbb{P}_\beta(E) - \mathbb{P}_{\beta'}(E)| = \sup_{E' \in \mathcal{X}} |\mathbb{P}_\beta(X \subseteq E') - \mathbb{P}_{\beta'}(X' \in E')|.$$

Note that since $\mathcal{X}$ is finite, this supremum is achieved for some $E' = E^*$.

With this setup, we have

$$\begin{aligned}\mathrm{TV}(\mathbb{P}_\beta, \mathbb{P}_{\beta'}) &= |\mathbb{P}_{0,\beta}(X \in E^*) - \mathbb{P}_{0,\infty}(X \in E^*)| \\ &\leq |\mathbb{E}\mathbf{1}\{X \in E^*\} - \mathbb{E}\mathbf{1}\{X' \in E^*\}| \\ &\leq \mathbb{E}|\mathbf{1}\{X \in E^*\} - \mathbf{1}\{X \in E^*\}| \\ &\leq \mathbb{P}(X \neq X').\end{aligned}$$

This completes the proof. □

Next, we have Pinsker's inequality, which may be found in Tsybakov (2009).

Lemma 5.5.3 (Pinsker's inequality)**.** *Let P and Q be two probability measures on the same discrete space. Then, we have*

$$TV(P, Q) \leq \sqrt{\frac{1}{2} KL(P, Q)}.$$

The next lemma is the log sum inequality, which is given as a simple consequence of Jensen's inequality on $f(x) = x \log x$ in Cover and Thomas (2012).

Lemma 5.5.4. *Let a_i and b_i be nonnegative numbers for $i = 1, \ldots, n$. Then, we have the inequality*

$$\left(\sum_{i=1}^n a_i\right) \log \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} \leq \sum_{i=1}^n a_i \log \frac{a_i}{b_i}.$$

Next, we have a standard concentration result, which may be found in Boucheron et al. (2013).

Lemma 5.5.5 (Bernstein's Inequality)**.** *Let X_i be centered random variables, i.e. $\mathbb{E}X_i = 0$, such that $|X_i| \leq M$ almost surely. Then for any $t > 0$, we have*

$$\mathbb{P}\left(\sum_{i=1}^n X_i > t\right) \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^n \mathbb{E}[X_i^2] + 2Mt/3}\right).$$

6. Numerical Details

In this Chapter, we discuss our HIV numerical analysis. We would like to accomplish two tasks: determine a confidence set for the parameter β and demonstrate how to provide an approximation of the confidence set for the underlying graph.

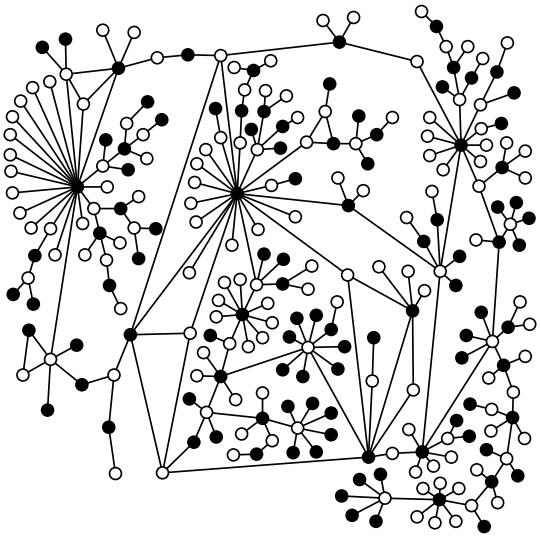

Figure 6.1 The HIV maxcut partition. The maximum cut consists of the sets of black and white vertices; the cut-set consists of the edges between the black and white vertices.

First, we comment on the computation of the upper bound $M_0 \leq 348.8$. We upper bounded each $\max_{P_{1:t-1}} W_t(P_{1:t-1})$ by the minimum of the maxcut value 258 and the sum of the degrees of the first $t-1$ infected vertices. The maxcut can be seen in Figure 6.1.

6.1 Confidence Set for β

Our first goal is to use the discretization of Definition 4.1.1 and the two-statistic test of Algorithm 5.3.2. For the confidence set algorithm, we set $\alpha = 0.2$, with $\alpha_0 = \alpha_1 = \alpha/2$. We set the parameters $\epsilon = 0.029$, $\xi = 0.001$, and $\gamma = 0.021$. Our two statistics are be W and $-W$, i.e. we wish to have two-sided confidence sets for the statistic W. For the discretization, we set $\delta = 0.02$, and we found have the upper bound $M_0 \leq 348.8$. Finally, we restrict to the set $\mathfrak{B} = [0, 10]$ in computing the confidence set. We plot the resulting p-values as a function of β in Figure 6.2.

If we consider a confidence interval formed by the smallest and largest values of β for which the p-value is greater than 0.05, the resulting interval is $[1.14435, 10]$. Note that an interval is a reasonable approximation when examining Figure 6.2. Thus, we have evidence for spreading on the HIV graph for β in this interval.

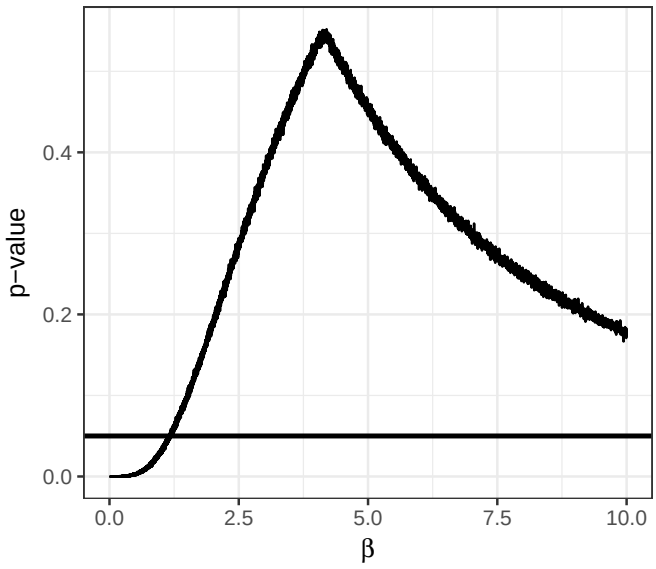

Figure 6.2 The p-values as a function of β for the HIV graph using the two-statistic test. The two statistics are W and $-W$, i.e. the p-value is small if W is too big or too small. Values of β above the horizontal line at 0.05 are in the 0.8-confidence set.

6.2 Toward a Confidence Set for $\mathcal{G}$

In order to build a confidence set for the entire graph structure, we can use the same simulation method on other graphs. For example, we could consider a graph that we call "graph 25" or $\mathcal{G}_{25}$, since it is the HIV graph with 25 randomly-chosen edges removed. A picture is available in Figure 6.3.

As discussed in Section 4.2.3, there are a number of different statistics that one might consider. Here, we consider two-sided tests for $W_{\mathcal{G}_{25}}$, $W_{\mathcal{G}_{\text{HIV}}}$, and $W_{\mathcal{G}_{25}} - W_{\mathcal{G}_{\text{HIV}}}$, i.e. the edges within statistic computed with respect to $\mathcal{G}_{25}$, the edges within statistic computed with respect to the HIV graph, and the difference of these statistics. The results are shown in Figure 6.4.

There are a number of things to observe. First, using the statistics $W_{\mathcal{G}_{25}}$ and $W_{\mathcal{G}_{\text{HIV}}}$ lead to confidence sets for β of approximately $[1.148, 10]$ and $[1.27347, 10]$ respectively. On the other hand, the difference $W_{\mathcal{G}_{25}} - W_{\mathcal{G}_{\text{HIV}}}$ leads to the confidence interval $[0, 10]$. This is unsurprising, since the $\mathcal{G}_{25}$ and $\mathcal{G}_{\text{HIV}}$ are not too different.

We can repeat this procedure on a different graph, $\mathcal{G}_{100}$, which has 100 edges removed from the HIV graph uniformly at random. An image may be found in Figure 6.5, and note that $\mathcal{G}_{100}$ is a subgraph of $\mathcal{G}_{25}$. The results of the analysis are shown in Figure 6.6.

The major difference for $\mathcal{G}_{100}$ is that no value of β is included in the confidence set for the difference statistic $W_{\mathcal{G}_{100}} - W_{\mathcal{G}_{\text{HIV}}}$. This is likely due to there being many edges in $\mathcal{G}_{\text{HIV}}$ between infected vertices that do not exist when simulating the distribution using $\mathcal{G}_{100}$. In

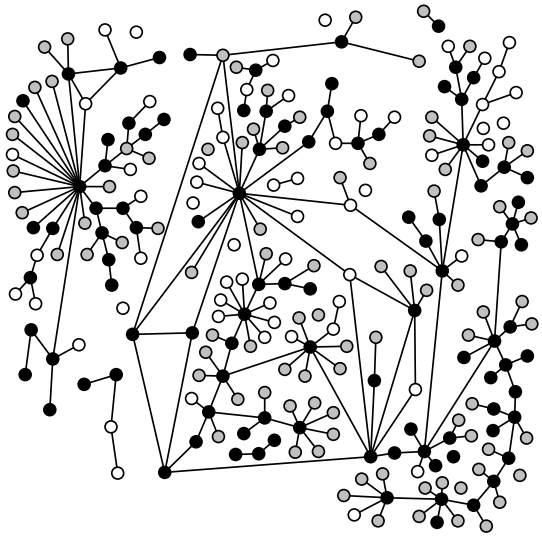

Figure 6.3 The graph $\mathcal{G}_{25}$ used in computing the joint confidence set of $(\mathcal{G}_{25}, \beta)$. This graph was formed from the HIV graph by removing 25 edges uniformly at random. As in Figure 2.1, black vertices are infected, white vertices are uninfected, and gray vertices are censored.

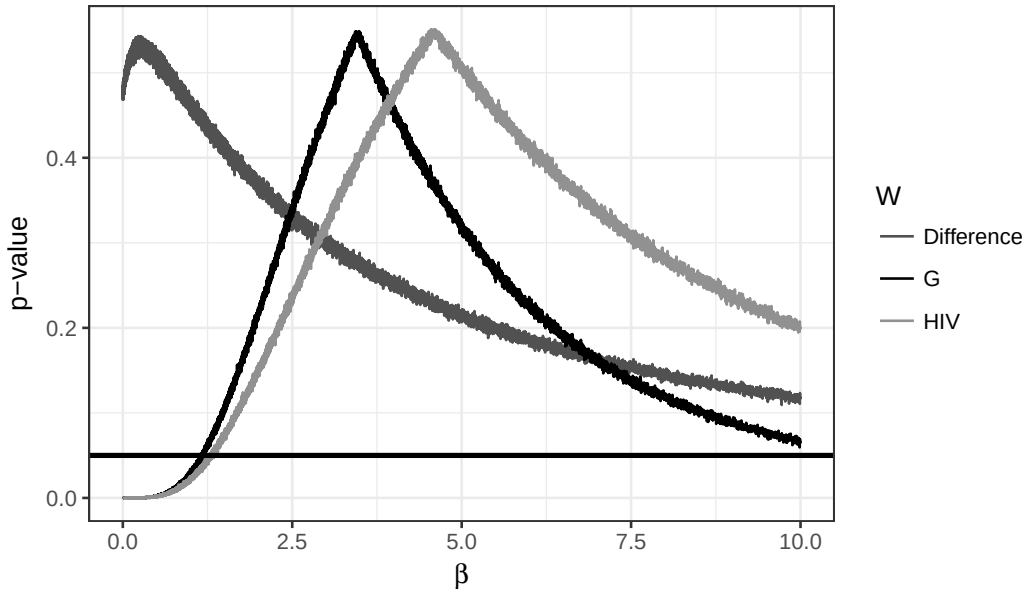

Figure 6.4 The p-values as a function of β for the HIV graph using the two-statistic test with different test statistics. The statistics are $W_{\mathcal{G}_{25}}$, $W_{\mathcal{G}_{\text{HIV}}}$, and $W_{\mathcal{G}_{25}} - W_{\mathcal{G}_{\text{HIV}}}$, and the p-values for these are given by the black line, the light gray line, and the dark graph line respectively.

short, for the values of β from $[0, 10]$, we have found a graph that is not in the confidence set of graphs given by the observed infection.

Thus, using this approach, we can test whether any given graph $\mathcal{G}$ belongs in the confidence

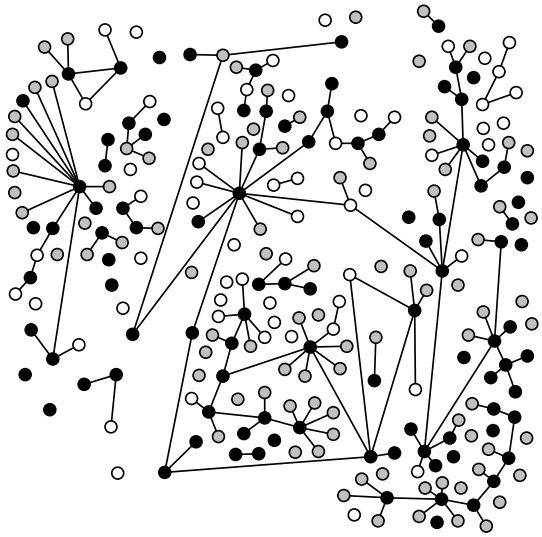

Figure 6.5 The graph $\mathcal{G}_{100}$ used in computing the joint confidence set of $(\mathcal{G}_{100}, \beta)$. This graph was formed from the HIV graph by removing 100 edges uniformly at random. As in Figure 2.1, black vertices are infected, white vertices are uninfected, and gray vertices are censored.

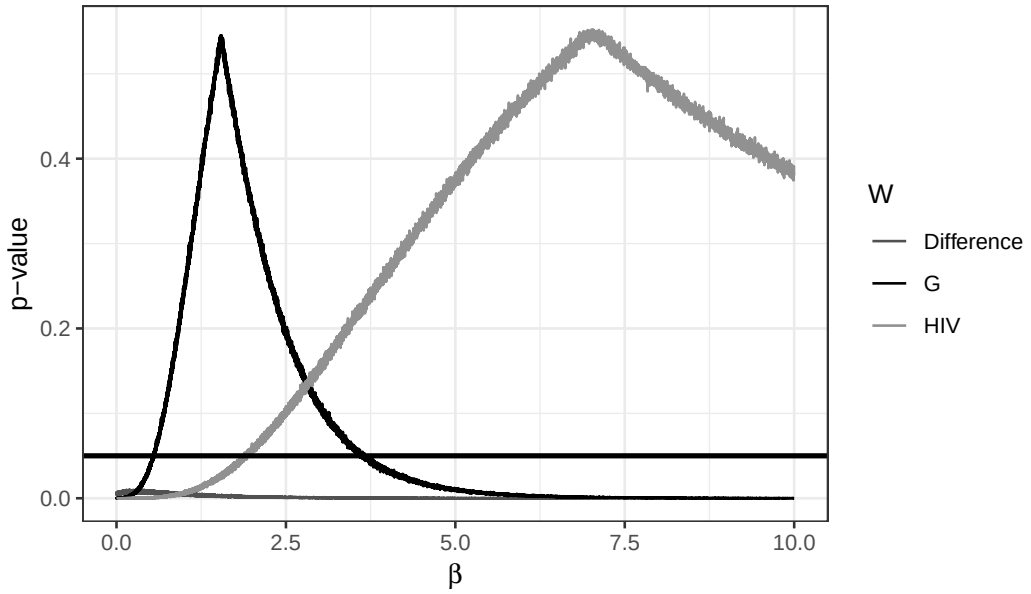

Figure 6.6 The p-values as a function of β for the HIV graph using the two-statistic test with different test statistics. The statistics are $W_{\mathcal{G}_{100}}$, $W_{\mathcal{G}_{\text{HIV}}}$, and $W_{\mathcal{G}_{100}} - W_{\mathcal{G}_{\text{HIV}}}$, and the p-values for these are given by the black line, the light gray line, and the dark graph line respectively.

set for this particular infection. This provides a very general method of graph testing at the expense of requiring many simulations.

Part III

Permutation

7. Theory

In this chapter, we cover the primary contribution of this dissertation: tests based on permutation. Note that we restrict our inquiry to the case where the edges are unweighted, or equivalently $w(u, v) = 1$.

7.1 Permutation-Invariant Statistics

A natural statistic to consider for the purpose of graph testing is the likelihood ratio. For the stochastic spreading model, the likelihood ratio is often difficult to compute and depends on β in a nontrivial manner, making the theoretical derivations somewhat challenging. Our main focus will be on a class of statistics that are invariant under a group of permutations, which allow us to perform permutation testing based on symmetries in the graph sets. We first introduce some terminology regarding permutations and group actions, and then introduce a class of invariant statistics that will be central to our analysis.

7.1.1 Permutations and Group Actions

Recall that a *graph automorphism* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is an element ϕ of the permutation group S_n such that $(u, v) \in \mathcal{E}$ if and only if $(\phi(u), \phi(v)) \in \mathcal{E}$. For simple hypotheses, we denote the automorphism groups of $\mathcal{G}_0$ and $\mathcal{G}_1$ by $\Pi_0 = \mathrm{Aut}(\mathcal{G}_0)$ and $\Pi_1 = \mathrm{Aut}(\mathcal{G}_1)$, respectively.

We also need to define the *action* of a permutation on vertices, graphs, and infections. The action of a permutation π on a vertex u is simply the image $\pi(u)$. This is easily extended to tuples and subsets of vertices by applying π to the underlying vertices. A specific example is the action on edges of the graph:

$$\pi\mathcal{E} = \{(\pi(u), \pi(v)) : (u, v) \in \mathcal{E}\}.$$

The action of π on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is then defined to be

$$\pi\mathcal{G} := (\pi\mathcal{V}, \pi\mathcal{E}) = (\mathcal{V}, \pi\mathcal{E}).$$

Another natural extension is to define the action of a set of permutations on a set of graphs:

$$\Pi\mathfrak{G} = \{\pi\mathcal{G} : \pi \in \Pi \text{ and } \mathcal{G} \in \mathfrak{G}\}.$$

If $\mathfrak{G}_i = S_n\{\mathcal{G}_i\}$, we say that hypothesis i corresponds to a hypothesis of a particular graph *topology*, since all node labelings are included in the set. We also define the action $\Pi\Theta_i = \Pi\mathfrak{G}_i \times A_i$. Finally, we define the action of a permutation π on an infection J:

$$\pi J := \left(J_{\pi^{-1}(1)}, \ldots, J_{\pi^{-1}(n)}\right).$$

In other words, the infection status of the image vertex $\pi(u)$ is the infection status of u under J.

7.1.2 Invariant Statistics

The theory presented in our paper applies to the following class of statistics:

Definition 7.1.1. *Suppose* Π *is a subgroup of* S_n*. A statistic* S *is* Π-invariant *if* $S(J) = S(\pi J)$ *for any* $J \in \mathbb{I}_{k,c}$ *and* $\pi \in \Pi$*.*

In our permutation test, we will compute the edges-within statistic with respect to the graph $\mathcal{G}_1$ appearing in the alternative hypothesis in the case of a simple test, so we reject H_0 when $W_1(J) := W_{\mathcal{G}_1}(J)$ exceeds a certain threshold. We derive the invariance of the statistic W under the permutation group $\Pi = \mathrm{Aut}(\mathcal{G})$ in Chapter 8.

7.2 Main Results

Our theoretical results are motivated by the following observation: when $\mathcal{G}_0$ is the empty graph, the coordinates of the infection vector $\mathcal{I}$ are exchangeable. Hence, we may conduct a valid permutation test based on any test statistic computed with respect to $\mathcal{I}$, where the rejection rule is given by the quantiles of the distribution of $\pi\mathcal{I}$, with $\pi \sim \mathrm{Uniform}(S_n)$. However, the permutation test remains valid in somewhat more general settings. In this section, we will assume the setting of "simple" hypothesis testing, where $\mathfrak{G}_0 = \{\mathcal{G}_0\}$ and $\mathfrak{G}_1 = \{\mathcal{G}_1\}$ are singleton sets. We will develop a sufficient condition, stated in terms of the interplay between the automorphism groups Π_0 and Π_1, which guarantees the validity of a permutation test applied to any Π_1-invariant statistic.

We let S denote a Π_1-invariant statistic. We also define

$$\Pi_{10} := \Pi_1\Pi_0 = \{\pi_1\pi_0 : \pi_i \in \Pi_i\}.$$

The following key theorem shows that the distribution of the test statistic is the same when applied to a random permutation of the infection vector, provided $\Pi_{10} = S_n$.

Theorem 7.2.1. *Suppose that for any* π_0 *in* Π_0*, the distribution of* $\mathcal{I}$ *satisfies*

$$\mathbb{P}_0(\mathcal{I} = J) = \mathbb{P}_0(\pi_0\mathcal{I} = J). \tag{7.1}$$

Let π *be drawn uniformly from* S_n*. If* $\Pi_{10} = S_n$*, the statistics* $S(\mathcal{I})$ *and* $S(\pi\mathcal{I})$ *have the same distribution under the null hypothesis.*

Showing that equation (7.1) holds for our models is straightforward, and we do this in Chapter 8. In particular, the condition $\Pi_{10} = S_n$ holds when $\mathcal{G}_0$ is the empty graph, since $\Pi_0 = S_n$ in that case. The next result shows that the condition described in Theorem 7.2.1 is sufficient to guarantee the success of a straightforward permutation test, described in Algorithm 7.2.1.

Theorem 7.2.2. *Suppose equation* (7.1) *holds and* $\Pi_{10} = S_n$*. The permutation test described in Algorithm 7.2.1 controls Type I error at level* α*.*

Algorithm 7.2.1: Permutation test (exact)

Input : Type I error tolerance $\alpha > 0$, observed infection vector $\mathcal{I}$

1 For each $\pi \in S_n$, compute the statistic $S(\pi\mathcal{I})$

2 Determine a threshold t_α such that

$$t_\alpha = \sup\left\{ t \in \text{supp}(S) : \frac{1}{n!} \sum_{\pi \in S_n} \mathbf{1}\{S(\pi\mathcal{I}) \geq t\} > \alpha \right\}$$

3 Reject H_0 if and only if $S(\mathcal{I}) > t_\alpha$

Remark 7.2.1. The proof of Theorem 7.2.2 critically leverages the property

$$S(\mathcal{I}) \overset{d}{=} S(\pi\mathcal{I}), \qquad \text{where } \pi \sim \text{Uniform}(S_n). \tag{7.2}$$

Note that this property would clearly hold in the case when the components of $\mathcal{I}$ are exchangeable, since we have $S(\mathcal{I}) \overset{d}{=} S(\pi\mathcal{I})$ for any fixed $\pi \in S_n$ in that case. Furthermore, under the "independent neighborhoods condition" of Milling et al. (2015), condition (7.2) holds, as well. However, the alternative graph $\mathcal{G}_1$ is randomly generated in such settings, so the statistic S is also random. We discuss this more precisely in Chapter 8.

For large values of n, it may be undesirable to compute $S(\pi\mathcal{I})$ for all permutations $\pi \in S_n$. Instead, we may approximate the rejection threshold t_α for the permutation test using Monte Carlo simulation, leading to Algorithm 7.2.2. As an immediate corollary to Theorem 7.2.2, Algorithm 7.2.2 is asymptotically accurate as $B \to \infty$.

Algorithm 7.2.2: Permutation test (approximate)

Input : Type I error tolerance $\alpha > 0$, integer $B \geq 1$, observed infection vector $\mathcal{I}$

1 Draw $\pi_1, \ldots, \pi_B \overset{i.i.d.}{\sim} \text{Uniform}(S_n)$ and compute the statistics $S(\pi_i\mathcal{I})$

2 Determine a threshold $\hat{t}_\alpha$ such that

$$\hat{t}_\alpha = \sup\left\{ t \in \text{supp}(S) : \frac{1}{B} \sum_{i=1}^{B} \mathbf{1}\{S(\pi_i\mathcal{I}) \geq t\} > \alpha \right\}$$

3 Reject H_0 if and only if $S(\mathcal{I}) > \hat{t}_\alpha$

Remark 7.2.2. Note that the permutation tests described in Algorithms 7.2.1 and 7.2.2 are very simple to execute and do not involve approximating the parameters λ or β in any way. Rather, the algorithms exploit differences in the symmetry structures of $\mathcal{G}_0$ and $\mathcal{G}_1$. As a caveat, the usefulness of the guarantee in Theorem 7.2.2 also depends on properties of the graphs $\mathcal{G}_0$ and $\mathcal{G}_1$ and their relationship to the test statistic S. In particular, if $S = W_1$ is the edges-within statistic and $\mathcal{G}_1$ is the empty graph, we always have $S = 0$. Thus, the threshold for the permutation test would be $t_\alpha = 0$, and the test would never reject H_0. Of course, this is a valid level-α test, but it has power 0. However, we prove rigorously in Section 7.3

below that our permutation test results in meaningful hypothesis testing procedures with reasonable risk bounds for a variety of interesting scenarios.

7.3 Examples and Risk Bounds

We now provide several examples to illustrate the use of Theorem 7.2.2. We focus on the cases when one graph is the star graph and the other is a vertex-transitive graph. In these cases, we are able to compute interpretable risk bounds for our permutation test; our calculations are valid under the stochastic spreading model, which we assume to be the setting for all the risk bounds computed in this section. We also assume a simple hypothesis testing scenario, where we only specify the parameter β for H_1 and let β be arbitrary for H_0. Also, we compute the risk for Algorithm 7.2.1, although a finite-simulation result could also be obtained for Algorithm 7.2.2 via a union bound argument.

The following corollary to Theorem 7.2.1 will be useful in our development. It implies that the risk incurred by testing $\mathcal{G}_0$ against $\mathcal{G}_1$ is exactly equal to the risk incurred by testing the empty graph against $\mathcal{G}_1$.

Corollary 7.3.1. *Suppose S is a Π_1-invariant statistic and $\Pi_{10} = S_n$. The risk of any test based on S is equal to the risk of the same test computed with respect to the null hypothesis H_0' involving the empty graph.*

In our examples, we analyze the simple case of a star graph versus a graph such that $\Pi_{10} = S_n$. These graphs are exactly the vertex transitive graphs. Recall the following definition (Godsil and Royle, 2013):

Definition 7.3.1. *A graph $\mathcal{G}$ is* vertex-transitive *if every pair of vertices is equivalent under some element of $\mathrm{Aut}(\mathcal{G})$; i.e. for any $u, v \in V(\mathcal{G})$, we have $\pi(u) = v$ for some $\pi \in \mathrm{Aut}(\mathcal{G})$.*

Basic examples of vertex-transitive graphs include the cycle graph and the toroidal grid.

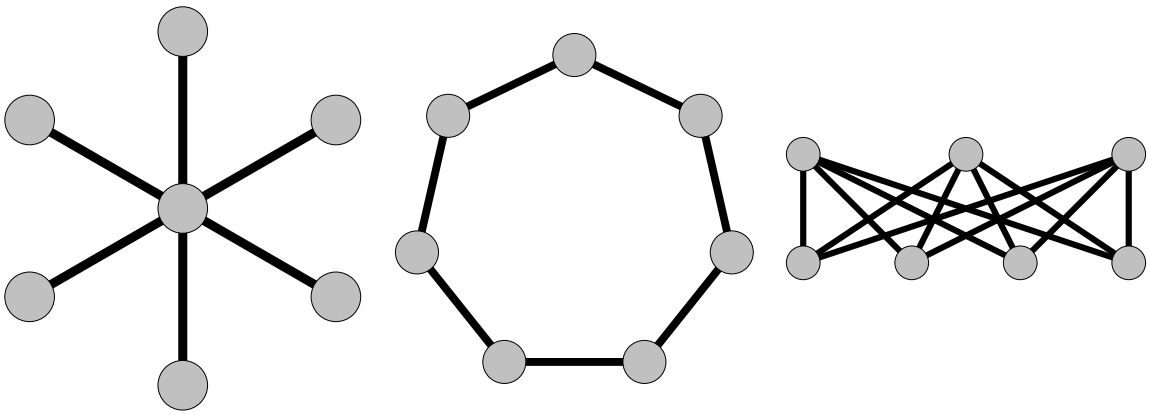

Figure 7.1 A 6-star, a 7-cycle, and the complete bipartite graph $K_{3,4}$. If $\mathcal{G}_0$ is the 6-star and $\mathcal{G}_1$ is the 7-cycle, then $\Pi = S_7$. If $\mathcal{G}_0$ is the 6-star and $\mathcal{G}_1 = K_{3,4}$, then $\Pi \neq S_7$.

7.3.1 The Star Graph as $\mathcal{G}_0$

First, let $\mathcal{G}_0$ be the star graph on n vertices. Without loss of generality, let vertex 1 be the center vertex. Note that $\Pi_0 \cong S_{n-1}$. In fact, $\Pi_{10} = S_n$ whenever Π_1 contains permutations mapping vertex 1 to any other vertex, which is equivalent to the definition of vertex transitivity. We summarize this observation in the following corollary.

Corollary 7.3.2. *Let $\mathcal{G}_0$ be the star graph and suppose $\mathcal{G}_1$ is vertex-transitive. Define the variable $\pi \sim$ Uniform(S_n), and suppose S is a Π_1-invariant statistic. Then $S(\mathcal{I}) \overset{d}{=} S(\pi\mathcal{I})$ under the null hypothesis, and the permutation test described in Algorithm 7.2.1 controls the Type I error at level α. A similar result holds when $\mathcal{G}_1$ is the star graph and $\mathcal{G}_0$ is vertex-transitive.*

We now turn to risk bounds. Before we present the bounds, we provide some motivation. When β is small, the two hypothesis spaces are close together, so the risk will be large; thus, we focus instead on the regime of moderate to large β. When the null hypothesis is true, the infection still looks uniformly random when viewing $\mathcal{G}_1$, as stated in Corollary 7.3.1. Using a standard concentration bound, we then may verify that t_α is $O(k^2/n + \sqrt{k})$, where the two terms correspond to the expected value and the variance of $W_1(\mathcal{I})$ under the null. When the alternative is true, the infected vertices should induce a connected subgraph of $\mathcal{G}_1$, with high probability. Thus, the expected value of $W_1(\mathcal{I})$ is $\Omega(k)$. Finally, we use another standard concentration bound to obtain an upper bound on the probability that $W_1(\mathcal{I}) \leq t_\alpha$ under the alternative.

Recall that all the vertices of a vertex-transitive graph have the same degree, which we denote by D. It is natural that the risk depends on D, since larger values of D allow for more variation in $W_1(\mathcal{I})$. Let $N_{t,\min}$ denote the minimum possible cut between the t infected vertices and $n-t$ uninfected vertices at time t. We define the function

$$H(\beta) = \prod_{m=1}^{k-1} \frac{\beta}{n - m + \beta N_{t,\min}}.$$

We also define a *cascade* on k vertices to be a surjective map $f : \mathcal{V} \to \{0, \ldots, k\}$, such that

(i) v is uninfected when $f(v) = 0$,

(ii) v is the i^{th} vertex infected when $f(v) = i$, and

(iii) if $f(v) = i$, then v must be adjacent to one of the first $i-1$ infected nodes.

Let $\mathcal{C}_k(u, v)$ denote the set of cascades on k vertices such that both u and v are infected, and let $C_k := \min_{(u,v)\in\mathcal{E}_1} |\mathcal{C}_k(u, v)|$. We have the following bound:

Proposition 7.3.1. *Suppose $\mathcal{G}_1$ is a connected vertex-transitive graph with degree D. Let $\psi_{W,\alpha}$ be the level-α permutation test based on the edges-within statistic W_1. Then*

$$R_{k,0}(\psi_{W,\alpha}, \beta) \leq \alpha + \exp\left\{ -\frac{2}{kD^2} \left(\frac{D}{2} C_k H(\beta) - \frac{Dk(k-1)}{2(n-1)} - \sqrt{\frac{kD^2}{2} \log \frac{1}{\alpha}} \right)^2 \right\}.$$

Remark 7.3.1. Note that $H(\beta)$ is increasing in β, so the risk bound in Proposition 7.3.1 decreases as β increases. This agrees with intuition, since higher values of β correspond to a higher chance that the infection propagates via edges rather than by random infections. Thus, the graphs $\mathcal{G}_0$ and $\mathcal{G}_1$ should be easier to distinguish.

For a cycle graph with $k < n/2$, we have the following result:

Corollary 7.3.3. *Let $\mathcal{G}_1$ be the n-cycle. Then $C_k = (k-1)2^{k-1}$ and $D = 2$, so*

$$R_{k,0}(\psi_{W,\alpha}, \beta) \leq \alpha + \exp\left\{ -\frac{1}{2k}\left((k-1)2^{k-1}\prod_{m=1}^{k-1}\frac{\beta}{n-m+2\beta} - \frac{k(k-1)}{n-1} - \sqrt{2k\log\left(\frac{1}{\alpha}\right)}\right)^2\right\}.$$

In particular, if $\alpha = \exp(-Ck/2)$ and $k/n + C \leq 1 - \epsilon$ for some $C > 0$ and $\epsilon > 0$, then there exist $C', C'' > 0$ such that

$$\lim_{\beta\to\infty} R_{k,0}(\psi_{W,\alpha}, \beta) \leq C' \exp(-C''k).$$

The last statement reveals that as $\beta \to \infty$, the risk will vanish for sufficiently large values of k. However, if the fraction k/n of infected nodes becomes too large, the two hypotheses are again difficult to distinguish.

7.3.2 The Star Graph as $\mathcal{G}_1$

We now consider the case when $\mathcal{G}_1$ is the star graph on n vertices. Again, let vertex 1 denote the center of the star. Perhaps unsurprisingly, it turns out that the maximum likelihood estimator and a test based on the edges-within statistic reduce to the same decision rule, depending on whether vertex 1 is included in the infected set.

Proposition 7.3.2. *Let $c = 0$. Suppose $\mathcal{G}_0$ is the empty graph. Maximum likelihood estimation is equivalent to the center indicator test $C = \mathbf{1}\{\mathcal{I}_1 = 1\}$, which is in turn equivalent to permutation testing at level α based on the edges-within statistic W_1, when $\alpha \geq k/n$.*

Risk bounds for hypothesis testing based on C are relatively easy to compute when $\mathcal{G}_0$ is the empty graph. Corollary 7.3.1 implies that such bounds hold for permutation testing when $\mathcal{G}_0$ is any vertex-transitive graph, from which we may derive the following result.

Proposition 7.3.3. *Suppose $\mathcal{G}_0$ is a vertex-transitive graph. The risk of the center indicator test on the star graph on n vertices satisfies the following bounds:*

$$R_{k,0}(P_C, \beta) \geq \frac{k}{n} + \exp\left(-\frac{k + \beta k(k-1)/2}{n-k}\right),$$

and

$$R_{k,0}(P_C, \beta) \leq \begin{cases} \frac{k}{n} + \exp\left(-\frac{k+\beta k(k-1)/2}{(n-k+1)+(k-1)\beta}\right), & \text{if } \beta \geq 1, \\ \frac{k}{n} + \exp\left(-\frac{k+\beta k(k-1)/2}{n}\right), & \text{if } \beta < 1. \end{cases}$$

Again, we can interpret the behavior of the risk bounds in terms of the fraction of infected vertices k/n. When β is fixed, the bound is $k/n + \exp(\Theta(-\beta k^2/n))$; thus, if we consider the size of the graph to be growing, we require $k/n \to 0$ and $k^2/n \to \infty$ in order to have vanishing risk. The first condition suggests that k cannot be large enough to randomly infect the center of the star under the null. The intuition for the latter condition is that under the alternative, each infected leaf attempts to infect the center of the star at successive time steps. This leads to at most $k(k-1)/2$ infection attempts, and when the strength of these attempts is large enough, the center is infected with high probability.

7.4 A Closer Look at Test Statistics

It is natural to wonder which of the invariant statistics leads to the best statistical test, or even whether it is reasonable to focus our attention on invariant statistics. We address the first question by showing a rough equivalence of likelihood ratio testing to tests based on the edges-within statistic. For the second question, we derive some results motivated by the Hunt-Stein theory of hypothesis testing.

7.4.1 Revisiting Likelihood: Likelihood Ratio and Edges-Within

As noted earlier, the edges-within statistic appears in the probability mass function of the 01-Ising model: $\mathcal{H}(J) = -W(J)$. Thus, a test based on a likelihood calculation may be equivalently expressed in terms of the edges-within statistic.

Turning to the stochastic spreading model, we now show that W_1 arises as the first-order coefficient in the series expansion of the likelihood with respect to β. This is somewhat reminiscent of the use of signed triangles in the random graph testing literature (Bubeck et al., 2016; Banerjee and Ma, 2017; Banerjee, 2018), which appears in the latter two cases as a first-order approximation to the asymptotic distribution of the log-likelihood. Furthermore, we will see that when considering the likelihood ratio of a test with a specific type of composite null hypothesis against the simple alternative $\mathfrak{G}_1 = \{\mathcal{G}_1\}$, the edges-within statistic W_1 also appears as the first-order coefficient of the likelihood ratio. These approximations are quite attractive, since as noted earlier, computing the likelihood ratio would require summing over $k!$ different infection paths and is itself intractable.

Before stating the main result, we introduce some additional notation. Let $L(\mathcal{G}, \beta; J)$ denote the likelihood of infection J under graph $\mathcal{G}$ and spreading parameter β. Let

$$R(\mathcal{G}_0, \mathcal{G}_1, \beta; J) = \frac{L(\mathcal{G}_1, \beta; J)}{L(\mathcal{G}_0, \beta; J)}$$

denote the likelihood ratio, and let $\mathcal{G}_{\text{empty}}$ denote the empty graph. Recall that N_t denotes the number of edges connecting infected vertices to uninfected vertices at time t.

Our main theorem shows the approximate equivalence between likelihood ratio tests and thresholding the edges-within statistic in the case of simple hypothesis testing when the null graph is empty.

Theorem 7.4.1. *Consider the hypothesis test of $\mathcal{G}_{empty}$ versus $\mathcal{G}_1$. For an uncensored $P \in \mathfrak{P}(J)$ in which $k + c'$ vertices are infected, define the function*

$$Q(P) = \sum_{t=1}^{k+c'} \frac{N_t}{n+1-t}.$$

If $c = 0$, we have

$$R(\mathcal{G}_{empty}, \mathcal{G}_1, \beta; J) = \Big(1 + \beta W_1(J) + O(\beta^2 W_1(J)^2)\Big)\left(1 - \frac{1}{k!}\sum_{P \in \mathfrak{P}(J)} \beta Q(P)\right).$$

If $c > 0$, we have

$$R(\mathcal{G}_{empty}, \mathcal{G}_1, \beta; J) = \sum_{J' \in \mathcal{U}(J)} D(\beta, k, c, |J'|)\Big(1 + \beta W_1(J') + O(\beta^2 W_1(J')^2)\Big) \times \left(1 - \frac{1}{|J'|!}\sum_{P \in \mathfrak{P}(J')} O(\beta Q(P))\right),$$

where

$$D(\beta, k, c, r) := \binom{n}{k\ \ c}\left(Z_1(\beta)\binom{n}{c}\binom{n}{r}\right)^{-1}.$$

Thus, Theorem 7.4.1 shows that when $c = 0$, the edges-within statistic $W_1(\mathcal{I})$ is the leading term of the likelihood ratio expanded as a function of β. When $c > 0$, the expression is somewhat more complicated, since the number of edges within the uncensored infection subgraph may differ depending on the statuses of censored vertices. The first-order term then equals the value of the edges-within statistic averaged over all possible infection vectors giving rise to the censored vector J.

Remark 7.4.1. Note that even if the likelihood ratio converges to its first-order approximation $1 + \beta W_1$ for some sequence of parameters, this does not ensure that W_1 will *always* lead to a useful test. For instance, in the case that $\mathcal{G}_1$ is a star graph, the condition $\beta W_1 \to 0$ and the conditions for asymptotically vanishing risk are incompatible. As discussed in Section 7.3.2, the latter conditions require $k/n \to 0$ and $\beta k^2/n \to \infty$. Since W_1 may take on the value $k - 1$, requiring that $\beta W_1, k/n \to 0$ would imply that $\beta k^2/n \to 0$, as well.

Finally, to obtain a test between two non-empty graphs, we use the simple relation

$$R(\mathcal{G}_0, \mathcal{G}_1, \beta; \mathcal{I}) = \frac{R(\mathcal{G}_{\text{empty}}, \mathcal{G}_1, \beta; \mathcal{I})}{R(\mathcal{G}_{\text{empty}}, \mathcal{G}_0, \beta; \mathcal{I})}.$$

This expression depends on both W_0 and W_1. However, cases exist where the expression only depends on W_1. For example, in the case of a composite null hypothesis involving testing a graph topology, we may derive the following theorem.

Theorem 7.4.2. *If $\mathfrak{G}_0 = S_n\mathfrak{G}_0$, then $\sup_{\theta\in\Theta_0} L(\theta; J)$ is constant over J in $\mathbb{I}_{k,c}$. Consequently, if we denote this constant by D', we have*

$$R(\mathfrak{G}_0, \mathcal{G}_1, \beta; J) = \frac{1}{D'} R(\mathcal{G}_{empty}, \mathcal{G}_1, \beta; J).$$

In conjunction with Theorem 7.4.1, Theorem 7.4.2 shows that we may again extract W_1 as the leading term in the expansion for the likelihood ratio. Another immediate corollary is that it is not possible to directly test between two unlabeled topologies, since the likelihoods for both hypotheses would be constant.

7.4.2 Hunt-Stein Theory for Graph Testing

When conducting hypothesis tests, it is natural to consider maximin tests, i.e. tests that maximize the minimum power over the space of alternative hypotheses. A standard way to do this for invariant tests is via the Hunt-Stein theorem (Lehmann and Romano, 2006). The goal of this section is to demonstrate that a version of the Hunt-Stein theorem also holds for the graph testing problem, further motivating our use of hypothesis testing via Π_1-invariant test statistics. The results in this subsection hold for general null and alternative parameter spaces $\Theta_0 = \mathfrak{G}_0 \times \mathbb{R}_+$ and $\Theta_1 = \mathfrak{G}_1 \times \mathbb{R}_+$. Let $\Theta := \Theta_0 \cup \Theta_1$.

Recall that a critical function φ outputs a value in $\{0, 1\}$ for each observed infection vector $\mathcal{I}$, corresponding to the selected hypothesis. A test based on φ is Π-invariant if $\varphi(\pi\mathcal{I}) = \varphi(\mathcal{I})$ for all $\pi \in \Pi$. Furthermore, the test is *maximin* at level α if

$$\max_{\theta\in\Theta_0} \mathbb{E}_\theta[\varphi(\mathcal{I})] \leq \alpha, \tag{7.3}$$

and the value of

$$\min_{\theta\in\Theta_1} \mathbb{E}_\theta[\varphi(\mathcal{I})]$$

is maximized among all tests φ' satisfying equation (7.3).

Analogous to canonical Hunt-Stein results, we will assume that both $\mathfrak{G}_0$ and $\mathfrak{G}_1$ are invariant under the same group of transformations, which in our setting is Π_1. A natural case where this condition is satisfied is when $\mathfrak{G}_1$ consists of a single graph $\mathcal{G}_1$ and $\mathfrak{G}_0$ consists of all permutations of a graph $\mathcal{G}_0$. For instance, suppose $\mathcal{G}_1$ is a cycle graph; if we wish to test a null hypothesis involving a star graph $\mathcal{G}_0$, we can define $\mathfrak{G}_0$ to include all permutations of $\mathcal{G}_0$ under Π_1, as well, which yields n stars with different center vertices.

Theorem 7.4.3 (Hunt-Stein for graph testing)**.** *Let Π be a group of transformations on $\mathcal{I}$, and let Θ_0, and Θ_1 be such that $\Pi\Theta_0 = \Theta_0$ and $\Pi\Theta_1 = \Theta_1$. If there exists a level-α test φ^* maximizing $\inf_{\theta\in\Theta} \mathbb{E}_\theta[\varphi(\mathcal{I})]$, then there also exists a Π-invariant test with this property, defined by*

$$\psi^*(J) = \frac{1}{|\Pi|} \sum_{\pi\in\Pi} \varphi^*(\pi J).$$

Thus, for composite tests in which Θ_0 and Θ_1 are both Π_1-invariant, whenever we can find a maximin critical function φ, we can also find a maximin Π_1-invariant critical function. Note, however, that one can do better than a Π_1-invariant test when $\mathfrak{G}_0$ is not Π_1-invariant, such as in the case of a star with a fixed center. We considered this example previously in Section 4.3.

7.5 Discussion

The results presented in this chapter suggest several avenues for further research. Although we only considered two infection models and closely examined the stochastic spreading model, the permutation testing framework could be extended to other infection models, as well. As an extension of Ising models, one could consider other classes of Markov random fields. Toward network modeling, such tests would also be helpful for testing network structure, such as Erdős-Renyi networks versus networks with block structure. We have left open the question of whether a version of Theorem 7.2.1 holds in more generality. It would also be useful to devise a more interpretable or verifiable sufficient condition for the validity of the permutation test. It might also be helpful to extend our results to settings where Π_0 and Π_1 are relatively small, in which case $\Pi_1\Pi_0 \neq S_n$. In fact, the proportion of graphs on n vertices with trivial automorphism groups tends to 1 as $n \to \infty$ (Godsil and Royle, 2013), so this would be the case for very large networks.

8. Technical Details

In this chapter, we provide additional theory and proofs for Chapter 7. Section 8.1 provides extensions to the theory, including (i) a partial converse to the main result of the paper; (ii) an extension to composite null and alternative hypotheses, (iii) an extension to the case of non-uniform censoring; (iv) a discussion of computing automorphism groups; (v) theory for multiple infection processes; (vi) a method for dealing with slight departures from the automorphism condition; and (vii) an extension to randomly-generated graphs.

Finally, we provide detailed proofs of all theoretical results stated in the paper. In Section 8.2, we provide proofs of our main theorems and propositions. In Section 8.3, we prove the corollaries. We provide proofs of technical lemmas in Section 8.4, and we conclude with a standard concentration result in Section 8.5.

8.1 Further Theoretical Results

This section contains additional theoretical results extending the main results of our paper. Proofs are contained in subsequent sections.

8.1.1 A Partial Converse

A natural question is whether the condition $\Pi_{10} = S_n$ is unnecessarily strong for guaranteeing the theory of our proposed permutation test. We now state and prove a partial converse to Theorem 7.2.1.

In particular, when $\mathcal{G}_0$ is a star graph, meaning a graph with $n-1$ edges connecting all nodes to a center node, the condition $\Pi_{10} = S_n$ is equivalent to the graph $\mathcal{G}_1$ being vertex-transitive. We have the following partial converse to this fact.

Theorem 8.1.1. *Let $\mathcal{G}_0$ be a star graph and $\mathcal{G}_1$ be a graph that is non-vertex-transitive. Let $\pi \sim Uniform(S_n)$. Then there exists a Π_1-invariant statistic S such that $S(\mathcal{I})$ and $S(\pi\mathcal{I})$ do not have the same distribution under H_0.*

In other words, if $\Pi_{10} \neq S_n$, it is possible to find a Π_1-invariant statistic that does *not* satisfy the sufficient condition that we use to derive the validity of the permutation test. We conjecture that a similar converse holds more broadly for general graphs $\mathcal{G}_0$; i.e. if $\Pi_1\Pi_0 \neq S_n$, a Π_1-invariant statistic S always exists such that $S(\mathcal{I}) \overset{d}{\neq} S(\pi\mathcal{I})$ under H_0.

8.1.2 Composite Null Hypotheses

Our next result concerns hypothesis tests involving a composite null hypothesis. In particular, note that the argument in Theorem 7.2.2 applies equally well to a composite null hypothesis, since we only require $\Pi_1\Pi_0 = S_n$ for all graphs $\mathcal{G}_0$ appearing in the null hypothesis. Hence,

the permutation test described in Algorithm 7.2.1 controls the Type I error at level α if all graphs $\mathcal{G}_0$ in the composite null hypothesis satisfy $\Pi_1\Pi_0 = S_n$.

The following result is an easy variant of Corollary 7.3.1, which we state without proof:

Corollary 8.1.1. *Suppose that S is Π_1-invariant statistic and $\mathcal{C}$ is a collection of graphs such that $\Pi_1 \operatorname{Aut}(\mathcal{G}_0) = S_n$ for all $\mathcal{G}_0 \in \mathcal{C}$. The risk of any test based on S is equal to the risk of the same test computed with respect to a single null hypothesis H_0' involving the empty graph.*

Such a result may be desirable in cases where there is uncertainty about the exact topology for a particular form of transmission, such as a water-borne illness, and we are testing against a very different transmission mechanism, such as a blood-borne disease. Testing star graph topologies could realistically arise in practice, for instance to network scientists who study networks from a game-theoretic or economic point of view, where star graphs naturally arise. On the other hand, the condition $\Pi_1\Pi_0 = S_n$ could be quite restrictive in other settings.

8.1.3 Composite Alternative Hypotheses

Suppose one wishes to conduct the following test:

$$H_0 : \mathcal{I} \sim \mathrm{SSM}(\mathcal{G}_0, \beta_0) \text{ for } \beta_0 \in [0, \infty]$$
$$H_1 : \mathcal{I} \sim \mathrm{SSM}(\mathcal{G}_1, \beta_1) \text{ for } \beta_1 > 0,\ \mathcal{G}_1 \in \mathfrak{G}_1.$$

For simplicity, we consider the two-graph alternative $\mathfrak{G}_1 = \{\mathcal{G}_1, \mathcal{G}_2\}$ and refer to the graph automorphism groups as Π_1 and Π_2. If $\mathcal{G}_1$ and $\mathcal{G}_2$ are not equivalent, i.e. there is no π_1 in Π_1 such that $\mathcal{G}_2 = \pi_1\mathcal{G}_1$, we consider the two-dimensional statistic (W_1, W_2), which is (Π_1, Π_2)-invariant. If $\Pi_1\Pi_0 = \Pi_2\Pi_0 = S_n$, both W_1 and W_2 have the same distribution under the null as if the true graph were empty; however, the two statistics could be correlated.

An exact permutation test is proposed in Algorithm 8.1.1, with theoretical guarantee in Theorem 8.1.2. The idea is to split the Type I error tolerance over the rejection regions of both edges-within statistics.

Algorithm 8.1.1: Permutation test for composite alternatives (exact)

Input : Type I error tolerance $\alpha > 0$, observed infection vector $\mathcal{I}$

1 For each $\pi \in S_n$, compute the statistic $(S_1(\pi\mathcal{I}), S_2(\pi\mathcal{I}))$

2 For $i = 1, 2$, determine a threshold $t_{\alpha,i}$ such that for $j < i$,

$$t_{\alpha,i} = \sup\left\{ t \in \mathrm{supp}(S) : \frac{1}{n!} \sum_{\pi \in S_n} \mathbf{1}\{S_i(\pi\mathcal{I}) \geq t \text{ and } S_j(\pi\mathcal{I}) \leq t_{\alpha,j}\} > \frac{\alpha}{2} \right\}$$

3 Reject H_0 if and only if $S_i(\mathcal{I}) > t_{\alpha,i}$ for either $i = 1$ or $i = 2$

Theorem 8.1.2. *Let $\pi \sim$ Uniform(S_n). Consider a hypothesis test of $\mathcal{G}_0$ versus the composite alternative $\mathfrak{G}_1 = \{\mathcal{G}_1, \mathcal{G}_2\}$. Let Π_0, Π_1, and Π_2 be the permutation groups of $\mathcal{G}_0$, $\mathcal{G}_1$, and*

$\mathcal{G}_2$ respectively, Let S_1 and S_2 be Π_1-invariant and Π_2-invariant statistics. If we have $\Pi_1\Pi_0 = \Pi_2\Pi_0 = S_n$, then $S_i(\mathcal{I})$ and $S_i(\pi\mathcal{I})$ have the same distribution under the null hypothesis for $i = 1, 2$. In particular, Algorithm 8.1.1 controls the Type I error at level α.

8.1.4 Conditioning on Censored Vertices

One important variant of the permutation test above involves conditioning on the censored vertices. This is crucial when censoring may not occur uniformly at random, but the stochastic spread still follows our model and occurs independently of the censoring. For a concrete example, consider the HIV graph.

Accordingly, we devise an alternative permutation testing procedure that conditions on the location of the censored nodes in the network. The natural analog of our permutation test is to simply build a histogram for the values of the test statistic under the subset of random permutations that fix the locations of the censored nodes. With a small abuse of notation, let S_{n-c} denote this set of $(n-c)!$ permutations. The exact and approximate permutation tests are described in Algorithms 8.1.2 and 8.1.3:

Algorithm 8.1.2: Permutation test (exact)

Input : Type I error tolerance $\alpha > 0$, observed infection vector $\mathcal{I}$

1 For each $\pi \in S_{n-c}$, compute the statistic $S(\pi\mathcal{I})$

2 Determine a threshold t_α such that

$$t_\alpha = \sup\left\{ t \in \text{supp}(S) : \frac{1}{(n-c)!} \sum_{\pi \in S_{n-c}} \mathbf{1}\{S(\pi\mathcal{I}) \geq t\} > \alpha \right\}$$

3 Reject H_0 if and only if $S(\mathcal{I}) > t_\alpha$

Algorithm 8.1.3: Permutation test (approximate)

Input : Type I error tolerance $\alpha > 0$, integer $B \geq 1$, observed infection vector $\mathcal{I}$

1 Draw $\pi_1, \ldots, \pi_B \overset{i.i.d.}{\sim} \text{Uniform}(S_{n-c})$ and compute the statistics $S(\pi_i\mathcal{I})$

2 Determine a threshold $\hat{t}_\alpha$ such that

$$\hat{t}_\alpha = \sup\left\{ t \in \text{supp}(S) : \frac{1}{B} \sum_{i=1}^{B} \mathbf{1}\{S(\pi_i\mathcal{I}) \geq t\} > \alpha \right\}$$

3 Reject H_0 if and only if $S(\mathcal{I}) > \hat{t}_\alpha$

The validity of the permutation tests is again stated in terms of automorphism groups of the null and alternative graphs, where we restrict our attention to automorphisms that fix the identity of the censored vertices. We define the set of possible infection vectors with k infected nodes and a fixed censoring pattern as

$$\mathbb{I}_{k,c}^{cc} = \{\mathcal{I} \in \{0,1\}^{n-c} \times \{\star\}^c : \mathcal{I} \text{ contains exactly } k \text{ 1's.}\},$$

where the "cc" stands for "conditioned on the censoring." If we let $\Pi_{0,cc}$ and $\Pi_{1,cc}$ denote the stabilizer subgroups of $\mathbb{I}^{cc}_{k,c}$ in Π_0 and Π_1; i.e.

$$\Pi_{i,cc} := \Big\{\pi \in \Pi_i : \pi \mathbb{I}^{cc}_{k,c} = \mathbb{I}^{cc}_{k,c}\Big\},$$

then permutations in $\Pi_{i,cc}$ act separately on the sets of censored and uncensored nodes. Thus, we may write $\Pi_{i,cc} = \Pi^u_{i,cc} \times \Pi^c_{i,cc}$ as a direct product of subgroups of Π_i acting on the uncensored and censored nodes. The analog of Theorems 7.2.1 and 7.2.2, guaranteeing the success of the permutation test, is stated in terms of the subgroup $\Pi^u_{10} = \Pi^u_{1,cc}\Pi^u_{0,cc}$.

Theorem 8.1.3. *Let $\pi \sim$ Uniform(S_{n-c}). If $\Pi^u_{10} = S_{n-c}$, then $S(\mathcal{I})$ and $S(\pi\mathcal{I})$ have the same distribution under the null hypothesis (conditioned on the identities of the censored nodes). In particular, the permutation test in Algorithm 8.1.2 controls Type I error at level α.*

The proof closely parallels that of Theorems 7.2.1 and 7.2.2, so we omit the details.

8.1.5 Computational Considerations

Here, we comment on the task of verifying the condition $\Pi_{10} = S_n$. As demonstrated in Section 7.3, it is sometimes possible to verify this condition analytically; however, we may need to check the condition computationally for a fixed pair of null and alternative graphs.

The computational complexity of computing the automorphism group of a graph is not known in general, and it is often studied as a reduction of the problem of determining whether two graphs are isomorphic (Lubiw, 1981). In the case of graphs of bounded degree, the problem is polynomial (Luks, 1982), and a number of algorithms have been proposed that test for nontrivial automorphisms, such as NAUTY (McKay, 1978), SAUCY Darga et al. (2004), and BLISS (Junttila and Kaski, 2007). Empirical evidence shows that these algorithms also perform reasonably well on moderately-sized graphs. Once Π_0 and Π_1 have been computed, we still need to verify that $\Pi = S_n$. One approach is to compute

$$|\Pi_{10}| = |\Pi_1\Pi_0| = \frac{|\Pi_1||\Pi_0|}{|\Pi_1 \cap \Pi_0|}, \tag{8.1}$$

and determine whether this expression is equal to $|S_n| = n!$. Equation (8.1) may be easily verified by considering cosets (Dummit and Foote, 2004).

8.1.6 Multiple Infection Spreads

Thus far, we have only discussed the case of observing a single infection spreading vector $\mathcal{I}$, but our results may easily be extended to the case of multiple spreads. Let $\mathcal{I}(1), \ldots, \mathcal{I}(m)$ be observation vectors from i.i.d. infection spreads on $\mathcal{G}_0$. Generalizing our earlier framework, we say that a statistic S is Π^m_1-*invariant* if

$$S(J(1), \ldots, J(m)) = S\Big(\pi^{(1)}J(1), \ldots, \pi^{(m)}J(m)\Big), \tag{8.2}$$

for any $J(1), \ldots, J(m) \in \mathbb{I}_{k,c}$ and any permutations $\pi^{(1)}, \ldots, \pi^{(m)} \in \Pi_1$. The proof of the following theorem is analogous to the proof of Theorem 7.2.1.

Theorem 8.1.4. *Let* $\pi^{(1)}, \dots, \pi^{(m)} \overset{i.i.d.}{\sim} \text{Uniform}(S_n)$. *If* $\Pi = S_n$, *then* $S(\mathcal{I}(1), \dots, \mathcal{I}(m))$ *and* $S\Big(\pi^{(1)}\mathcal{I}(1), \dots, \pi^{(m)}\mathcal{I}(m)\Big)$ *have the same distribution under* H_0.

Note that when multiple spreads are observed on the same graph, one natural approach is to infer the edges of the graph using an appropriate estimation procedure (Gomez-Rodriguez et al., 2016; Netrapalli and Sanghavi, 2012). On the other hand, Theorem 8.1.4 shows that a permutation test may also be employed in a hypothesis testing framework. The description of the permutation test is identical to Algorithm 7.2.1, except the statistic $S(\mathcal{I})$ is replaced by $S\Big(\pi^{(1)}\mathcal{I}(1), \dots, \pi^{(m)}\mathcal{I}(m)\Big)$, and the average is taken over all $(n!)^m$ possible choices of $\Big(\pi^{(1)}, \dots, \pi^{(m)}\Big)$.

A special case of a Π_1^m-invariant statistic is the average of all edges-within statistics.

$$\overline{W}\Big(\pi^{(1)}\mathcal{I}(1), \dots, \pi^{(m)}\mathcal{I}(m)\Big) := \frac{1}{m}\sum_{i=1}^{m} W_{\mathcal{G}_1}\Big(\pi^{(i)}\mathcal{I}(i)\Big). \tag{8.3}$$

The dependence on m leads to an exponential reduction in the Type II error, due to the concentration of the empirical average of m samples of the edges-within statistic.

Proposition 8.1.1. *Suppose* $\mathcal{G}_0$ *is the star graph and* $\mathcal{G}_1$ *is a connected vertex-transitive graph of degree* D. *Let* $\psi_{\overline{W},\alpha}$ *be the level* α *permutation test based on the average edges-within statistic* (8.3). *The risk of this test is bounded by*

$$R_{k,0}(\psi_{\overline{W},\alpha}, \beta) \le \alpha + \exp\left\{-\frac{2m}{kD^2}\left(\frac{D}{2}C_k H(\beta) - \frac{Dk(k-1)}{2(n-1)} - \sqrt{\frac{kD^2}{2m}\log\frac{1}{\alpha}}\right)^2\right\}.$$

We may also obtain analogs of Propositions 7.3.2 and 7.3.3. The MLE rejects H_0 when the average center indicator statistic $\overline{C}$ exceeds a threshold. We have the following result.

Proposition 8.1.2. *Suppose* $\mathcal{G}_0$ *is a vertex-transitive graph and* $\mathcal{G}_1$ *is the star graph. Let* $\psi_{\overline{C},\alpha}$ *be the level* α *permutation test based on the average center indicator statistic. Let*

$$p_{k,0}(\beta) := \begin{cases} 1 - \exp\Big(-\frac{k+\beta k(k-1)/2}{(n-k+1)+(k-1)\beta}\Big) & \text{if } \beta \ge 1 \\ 1 - \exp\Big(-\frac{k+\beta k(k-1)/2}{n}\Big) & \text{if } \beta < 1. \end{cases}$$

Then we have the risk bound

$$R_{k,0}(\psi_{\overline{C},\alpha}, \beta) \le \alpha + \exp\left(-2m\left(\frac{k}{n} + \sqrt{\frac{1}{2m}\log\frac{1}{\alpha}} - p_{k,0}(\beta)\right)^2\right).$$

8.1.7 Relaxing the Choice of Alternative Graph

If $\Pi_1\Pi_0 \neq S_n$, then $S(\mathcal{I})$ and $S(\pi\mathcal{I})$ may not have the same distribution under H_0, undermining the theory behind Algorithm 7.2.1. Nonetheless, it may be worthwhile to consider an alternative statistic S' that is Π_1'-invariant for a subset $\Pi_1' \subseteq S_n$ such that $\Pi_1'\Pi_0 = S_n$. A simple modification of the proof of Theorem 7.2.1 furnishes the following result.

Theorem 8.1.5. *Let $\Pi_0 = \mathrm{Aut}(\mathcal{G}_0)$, and let Π_1' be a subset of S_n such that $\Pi_1'\Pi_0 = S_n$. Let $\pi \sim$ Uniform(S_n). If S' is a Π_1'-invariant statistic; i.e., $S'(J) = S'(\pi_1' J)$ for all π_1' in Π_1' and J in $\mathbb{I}_{k,c}$, then $S'(\mathcal{I})$ and $S'(\pi\mathcal{I})$ have the same distribution under the null hypothesis.*

In particular, Theorem 8.1.5 establishes that the permutation test in Algorithm 7.2.1 controls the Type I error at level α for any Π_1'-invariant statistic. Theorem 8.1.5 may provide useful guarantees when $\mathcal{G}_1$ is close to having Π_1' as its automorphism group. For instance, we may have $\Pi_1' = \mathrm{Aut}(\mathcal{G}_1')$ for some graph $\mathcal{G}_1'$ that is only a slight modification of $\mathcal{G}_1$. Consider the following example: Let $\mathcal{G}_0$ be the star graph on n vertices, and let $\mathcal{G}_1$ be the line graph L_n on n vertices. Clearly, L_n is not vertex-transitive, so $\Pi_1\Pi_0 \neq S_n$. On the other hand, for large n, the line graph is almost the cycle graph, which we denote by C_n. Let W_{L_n} and W_{C_n} be the edges-within statistic on the line graph and the cycle graph. Then

$$W_{C_n}(J) = W_{L_n}(J) + \mathbf{1}\{J_n = J_1 = 1\}.$$

As a result, these statistics are quite similar, so the risk of a permutation test based on the statistic W_{C_n} is also similar. We have the following result.

Corollary 8.1.2. *Let $\mathcal{G}_1 = L_n$, and let $\psi_{W,\alpha}$ be the level α permutation test based on W_{C_n}. Suppose $k < n/2$. Then*

$$R_{k,0}(\psi_{W,\alpha}, \beta) \leq \alpha + \exp\left\{ -\frac{1}{2k}\left(\frac{1}{n}\cdot(k-1)2^{k-1}\cdot\frac{n-k+1}{n} \right.\right. \\ \left.\left. \cdot \prod_{m=1}^{k-1} \frac{\beta}{n-m+2\beta} - \frac{k(k-1)}{n-1} - \sqrt{2k\log\left(\frac{1}{\alpha}\right)}\right)^2\right\}.$$

Compared with Corollary 7.3.3, the expression in Corollary 8.1.2 only contains an additional factor of $(n-k+1)/n$ in the first term of the exponent.

8.1.8 Random Graphs

For our final extension, we turn to random graphs. We first introduce some additional notation.

Throughout this section, consider $\mathcal{G}_0$ and $\mathcal{G}_1$ to be random variables, and let Π_0'' and Π_1'' be the subgroups of permutations such that for any fixed graph G and permutation π_i in Π_i'', we have

$$\mathbb{P}(\mathcal{G}_i = G) = \mathbb{P}(\pi_i\mathcal{G}_i = G). \tag{8.4}$$

As an example, an Erdős-Renyi random graph satisfies equation (8.4) with $\Pi_i'' = S_n$. Let $S(\mathcal{I}, \mathcal{G}_1)$ be a statistic, where we explicitly include the graph dependence. We are most interested in $W(\mathcal{I}, \mathcal{G}_1)$, which we use to denote the usual edges-within statistic on the random graph $\mathcal{G}_1$. We have the following lemma.

Lemma 8.1.1. *Suppose that under H_0, the random variables $\mathcal{I}$ and $\mathcal{G}_1$ satisfy*

$$\mathbb{P}(S(\mathcal{I}, \mathcal{G}_1) \in \mathcal{B}) = \mathbb{P}(S(\pi\mathcal{I}, \mathcal{G}_1) \in \mathcal{B}), \tag{8.5}$$

for all π in S_n. Then we have

$$\mathbb{P}(S(\mathcal{I}, \mathcal{G}_1) \in \mathcal{B}) = \frac{1}{n!} \sum_{\pi \in S_n} \mathbb{P}(S(\pi\mathcal{I}, \mathcal{G}_1) \in \mathcal{B}). \tag{8.6}$$

The proof is obtained by averaging over S_n.

Remark 8.1.1. The condition in equation (8.5) is satisfied in a number of cases. One example is the following: if $\mathcal{G}_0$ and $\mathcal{G}_1$ are drawn independently and the former is an Erdős-Renyi random graph, we have

$$\begin{aligned}
\mathbb{P}((\mathcal{I}, \mathcal{G}_1) = (J, G)) &= \mathbb{P}(\mathcal{I} = J)\mathbb{P}(\mathcal{G}_1 = G) \\
&= \mathbb{P}(\pi\mathcal{I} = J)\mathbb{P}(\mathcal{G}_1 = G) \\
&= \mathbb{P}((\pi\mathcal{I}, \mathcal{G}_1) = (J, G)),
\end{aligned}$$

for any π in S_n. The first and last equalities are due to independence, and the second is the result of the S_n-invariance of the Erdős-Renyi random graph.

Another example satisfying equation (8.5) is the case where $S = W$ and the topology of the alternative $\mathcal{G}_1$ is fixed, but the labeling of the vertices is uniformly random, such as in Milling et al. (2015). In this case, we have

$$\mathbb{P}(W(\mathcal{I}, \mathcal{G}_1) \in \mathcal{B}) = \mathbb{P}(W(\mathcal{I}, \pi\mathcal{G}_1) \in \mathcal{B}) = \mathbb{P}(W(\pi\mathcal{I}, \mathcal{G}_1) \in \mathcal{B}),$$

where we have used the fact that $\Pi_1'' = S_n$ and the equality $W(\mathcal{I}, \pi\mathcal{G}_1) = W(\pi\mathcal{I}, \mathcal{G}_1)$. Additionally, we have assumed the independence of $\mathcal{G}_0$ and $\mathcal{G}_1$.

Turning to the conclusion of Lemma 8.1.1, note that equation (8.6) implies that the permutation test controls the Type I error on average with respect to draws of $\mathcal{G}_1$. In the special case where the topology of $\mathcal{G}_1$ is fixed—or alternatively, $\mathcal{G}_1$ is deterministic up to a permutation—equation (8.6) implies that the permutation test does control the Type I error. To see this, suppose that $\mathcal{G}_1$ is chosen from $\mathfrak{G} = S_n\{G^\star\}$ for some fixed $G^\star$, and let S

be a statistic that is invariant in the sense that $S(\mathcal{I}, G) = S(\pi\mathcal{I}, \pi G)$. We have

$$\begin{aligned}
\mathbb{P}(S(\mathcal{I}, \mathcal{G}_1) \in \mathcal{B}) &= \frac{1}{n!} \sum_{\pi \in S_n} \mathbb{P}(S(\pi\mathcal{I}, \mathcal{G}_1) \in \mathcal{B}) \\
&= \frac{1}{n!} \sum_{\pi \in S_n} \sum_{G \in \mathfrak{G}} \mathbb{P}(\mathcal{G}_1 = G) \mathbb{P}(S(\pi\mathcal{I}, G) \in \mathcal{B}) \\
&= \frac{1}{n!} \sum_{\pi \in S_n} \sum_{G \in \mathfrak{G}} \mathbb{P}(\mathcal{G}_1 = G) \mathbb{P}(S(\pi\mathcal{I}, \pi' G^\star) \in \mathcal{B}) \\
&= \frac{1}{n!} \sum_{\pi'' \in S_n} \sum_{G \in \mathfrak{G}} \mathbb{P}(\mathcal{G}_1 = G) \mathbb{P}(S(\pi''\mathcal{I}, G^\star) \in \mathcal{B}) \\
&= \frac{1}{n!} \sum_{\pi'' \in S_n} \mathbb{P}(S(\pi''\mathcal{I}, G^\star) \in \mathcal{B}) \sum_{G \in \mathfrak{G}} \mathbb{P}(\mathcal{G}_1 = G) \\
&= \frac{1}{n!} \sum_{\pi'' \in S_n} \mathbb{P}(S(\pi''\mathcal{I}, G^\star) \in \mathcal{B}).
\end{aligned}$$

Note that the second equality is by independence, the third equality is by the deterministic topology, and the fourth equality is by the invariance of S described above. Thus, the permutation test indeed controls the Type I error in this case.

8.2 Proofs

In this section, we provide proofs of our main results and propositions.

8.2.1 Proofs of Main Results

Proof of Theorem 7.2.1. Note that under H_0, and for any set $\mathcal{B}$, we have

$$\mathbb{P}_0(\mathcal{I} \in \mathcal{B}) = \mathbb{P}_0(\pi\mathcal{I} \in \mathcal{B}) = \frac{1}{|\Pi_0|} \sum_{\pi_0 \in \Pi_0} \mathbb{P}_0(\pi_0\mathcal{I} \in \mathcal{B}),$$

where the first equality is by assumption and the second is by averaging. The first equality is proven in detail for our models in Section 8.2.2. In particular,

$$\begin{aligned}
\mathbb{P}_0(S(\mathcal{I}) \leq t) &= \mathbb{P}_0\Big(\mathcal{I} \in S^{-1}(-\infty, t]\Big) \\
&= \frac{1}{|\Pi_0|} \sum_{\pi_0 \in \Pi_0} \mathbb{P}_0\Big(\pi_0\mathcal{I} \in S^{-1}(-\infty, t]\Big) \\
&= \frac{1}{|\Pi_0|} \sum_{\pi_0 \in \Pi_0} \mathbb{P}_0(S(\pi_0\mathcal{I}) \leq t).
\end{aligned}$$

Now let $\{g_1, \ldots, g_m\}$ denote coset representatives for Π_0 in the permutation group S_n, so $\{g_1\Pi_0, \ldots, g_m\Pi_0\}$ partitions S_n and has cardinality

$$m = \frac{|S_n|}{|\Pi_0|} = \frac{n!}{|\Pi_0|}.$$

By assumption, we may choose $g_i \in \Pi_1$ for each $1 \le i \le m$. Since S is a Π_1-invariant statistic, this means $S(g_i\pi_0\mathcal{I}) = S(\pi_0\mathcal{I})$ for each i. Hence, we have

$$\mathbb{P}_0(S(\mathcal{I}) \le t) = \frac{1}{|\Pi_0|}\sum_{\pi_0\in\Pi_0}\frac{1}{m}\sum_{1\le i\le m}\mathbb{P}_0(S(g_i\pi_0\mathcal{I}) \le t) = \frac{1}{n!}\sum_{\pi\in S_n}\mathbb{P}_0(S(\pi\mathcal{I}) \le t),$$

implying that $S(\mathcal{I})$ and $S(\pi\mathcal{I})$ have the same distribution when $\pi \sim \text{Uniform}(S_n)$. □

Proof of Theorem 7.2.2. By Theorem 7.2.1, we have

$$\mathbb{P}_0(S(\mathcal{I}) \le t) = \frac{1}{n!}\sum_{\pi\in S_n}\mathbb{P}_0(S(\pi\mathcal{I}) \le t),$$

for each $t \in \mathbb{R}$. Conditioning on the infection vector $\mathcal{I}$, we obtain

$$\begin{aligned}\mathbb{P}_0(S(\mathcal{I}) \le t) &= \sum_{\mathcal{I}\in\mathbb{I}_{k,c}}\mathbb{P}_0(\mathcal{I})\frac{1}{n!}\sum_{\pi\in S_n}\mathbb{P}_0\Big(S(\pi\mathcal{I}) \le t \,\Big|\, \mathcal{I}\Big)\\ &= \frac{1}{n!}\sum_{\mathcal{I}\in\mathbb{I}_{k,c}}\sum_{\pi\in S_n}\mathbb{P}_0(\mathcal{I})\mathbb{P}_0\Big(S(\pi\mathcal{I}) \le t \,\Big|\, \mathcal{I}\Big).\end{aligned}$$

Fix some $J \in \mathbb{I}_{k,c}$. For an infection $\mathcal{I}$, let $\pi_{J,\mathcal{I}}$ be a permutation mapping $\mathcal{I}$ to J. Then $\mathcal{I} = \pi_{J,\mathcal{I}}^{-1}J$, so $\pi\mathcal{I} = \pi\pi_{J,\mathcal{I}}^{-1}J$, and

$$\begin{aligned}\mathbb{P}_0(S(\mathcal{I}) \le t) &= \frac{1}{n!}\sum_{\mathcal{I}\in\mathbb{I}_{k,c}}\sum_{\pi\in S_n}\mathbb{P}_0(\mathcal{I})\mathbb{P}_0\Big(S\Big(\pi\pi_{J,\mathcal{I}}^{-1}J\Big) \le t \,\Big|\, \mathcal{I}\Big)\\ &= \frac{1}{n!}\sum_{\mathcal{I}\in\mathbb{I}_{k,c}}\mathbb{P}_0(\mathcal{I})\sum_{\pi'\in S_n}\mathbb{P}_0\Big(S(\pi'J) \le t \,\Big|\, \mathcal{I}\Big)\\ &= \frac{1}{n!}\sum_{\mathcal{I}\in\mathbb{I}_{k,c}}\mathbb{P}_0(\mathcal{I})\sum_{\pi'\in S_n}\mathbb{P}_0(S(\pi'J) \le t),\end{aligned}$$

where we denote $\pi' = \pi\pi_{J,\mathcal{I}}^{-1}$. In the last step, the conditioning on $\mathcal{I}$ becomes irrelevant because we sum over all permutations in S_n. Note that

$$\mathbb{P}_0(S(\pi'J) \le t) = \mathbf{1}\{S(\pi'J) \le t\},$$

since all quantities are deterministic. Hence,

$$\mathbb{P}_0(S(\mathcal{I}) \le t) = \frac{1}{n!}\sum_{\mathcal{I}\in\mathbb{I}_{k,c}}\mathbb{P}_0(\mathcal{I})\sum_{\pi'\in S_n}\mathbf{1}\{S(\pi'J) \le t\} = \frac{1}{n!}\sum_{\pi'\in S_n}\mathbf{1}\{S(\pi'J) \le t\}.$$

This justifies the permutation test described in Algorithm 7.2.1: The threshold t_α to bound the Type I error at level α may be computed explicitly from computing the appropriate quantile of S with respect to all $n!$ permutations of J. □

Proof of Theorem 8.1.1. Let $\mathcal{O}_{\Pi_1}(1)$ denote the orbit of vertex 1 in $\Pi_1 = \mathrm{Aut}(G_1)$, and note that by assumption, $|\mathcal{O}_{\Pi_1}(1)| < n$. Consider the statistic

$$S(J) = \sum_{v \in \mathcal{O}_{\Pi_1}(1)} 1\{J_v = 1\},$$

which counts the number of infected vertices in $\mathcal{O}_{\Pi_1}(1)$. Note that S is clearly Π_1-invariant. We claim that $S(\mathcal{I})$ and $S(\pi\mathcal{I})$ do not have the same distribution under H_0, when $\pi \sim \mathrm{Uniform}(S_n)$.

Let $G = \{g_1, \ldots, g_a\} \subseteq \Pi_0$ be a set consisting of coset representatives such that the collection of sets $\{\Pi_1 g_1, \ldots, \Pi_1 g_a\}$ is a partition of $\Pi_1\Pi_0$, and let $H = \{h_1, \ldots, h_b\} \subseteq S_n \backslash \Pi_0$ be representatives of the remaining cosets of Π_1 in S_n. For any observation vector J, we have

$$\begin{aligned}
\mathbb{P}_0(S(\pi\mathcal{I}) = S(J)) &= \frac{1}{n!} \sum_{\pi_1 \in \Pi_1} \sum_{g \in G \cup H} \mathbb{P}_0(S(\pi_1 g\mathcal{I}) = S(J)) \\
&= \frac{|\Pi_1|}{n!} \sum_{g \in G \cup H} \mathbb{P}_0(S(g\mathcal{I}) = S(J)) \\
&= \frac{1}{a+b} \sum_{g \in G \cup H} \mathbb{P}_0(S(g\mathcal{I}) = S(J)),
\end{aligned}$$

where the first equality uses the fact that $\{\Pi_1 g_1, \ldots, \Pi_1 g_a\} \cup \{\Pi_1 h_1, \ldots, \Pi h_b\}$ is a partition of S_n, and the second equality uses the fact that S is Π_1-invariant.

By symmetry of the spreading process on $\mathcal{G}_0$, we have

$$\mathbb{P}_0(S(g\mathcal{I}) = S(J)) = \mathbb{P}_0(S(\mathcal{I}) = S(J)), \qquad \forall g \in G.$$

Hence,

$$\mathbb{P}_0(S(\pi\mathcal{I}) = S(J)) = \frac{a}{a+b}\mathbb{P}_0(S(\mathcal{I}) = S(J)) + \frac{1}{a+b} \sum_{h \in H} \mathbb{P}_0(S(h\mathcal{I}) = S(J)).$$

We will demonstrate a choice of J for which

$$\sum_{h \in H} \mathbb{P}_0(S(h\mathcal{I}) = S(J)) \neq b\, \mathbb{P}_0(S(\mathcal{I}) = S(J)),$$

implying that

$$\mathbb{P}_0(S(\pi\mathcal{I}) = S(J)) \neq \mathbb{P}_0(S(\mathcal{I}) = S(J)),$$

so $S(\pi\mathcal{I})$ and $S(\mathcal{I})$ cannot have the same distribution.

Let $J \in \mathbb{I}_{k,c}$ be a vector such that $S(J) = m$, for some m to be specified later. Then $\mathbb{P}_0(S(\mathcal{I}) = S(J))$ is the probability that exactly m vertices in $\mathcal{O}_{\Pi_1}(1)$ are infected. On the other hand, for a fixed $g \in S_n$, the quantity $S(g\mathcal{I})$ counts the number of infected vertices in $g^{-1}(\mathcal{O}_{\Pi_1}(1))$. Again using symmetry of the spreading process on $\mathcal{G}_0$, we have

$$\mathbb{P}_0(S(g\mathcal{I}) = m) = \mathbb{P}_0(S(\mathcal{I}) = m),$$

whenever $1 \in g^{-1}(\mathcal{O}_{\Pi_1}(1))$. However, when $1 \notin g^{-1}(\mathcal{O}_{\Pi_1}(1))$, we have

$$\mathbb{P}_0(S(g\mathcal{I}) = m) < \mathbb{P}_0(S(\mathcal{I}) = m),$$

for some m, since the center of the star is more likely to be infected than any of the leaves. Finally, note that $h_j(1) \notin \mathcal{O}_{\Pi_1}(1)$ for some j. Indeed, if $h_i(1) \in \mathcal{O}_{\Pi_1}(1)$ for all $1 \leq i \leq b$, we would have $\pi_1 h_i(1) \in \mathcal{O}_{\Pi_1}(1)$ for all $\pi_1 \in \Pi_1$, contradicting the fact that the cosets cover the entire space S_n. Thus, we have $1 \notin h_j^{-1}(\mathcal{O}_{\Pi_1}(1))$, implying that

$$\mathbb{P}_0(S(h_j\mathcal{I}) = m) < \mathbb{P}_0(S(\mathcal{I}) = m),$$

and in particular,

$$\sum_{i=1}^{b} \mathbb{P}_0(S(h_i\mathcal{I}) = S(J)) < b\, \mathbb{P}_0(S(\mathcal{I}) = S(J)).$$

This completes the proof. □

8.2.2 Proofs of Invariances

In this section, we derive the following key lemma concerning permutation invariance of infection statistics under the graph automorphism group. This lemma is leveraged in the proof of Theorem 7.2.1. We first outline the proof of the lemma, and then provide the proofs of several supporting lemmas.

Lemma 8.2.1. *Let $\Pi_0 = \mathrm{Aut}(\mathcal{G}_0)$. For any π_0 in Π_0, we have*

$$\mathbb{P}_0(\mathcal{I} = J) = \mathbb{P}_0(\pi_0\mathcal{I} = J) = \frac{1}{|\Pi_0|} \sum_{\pi \in \Pi_0} \mathbb{P}_0(\pi\mathcal{I} = J),$$

under both the stochastic spreading and conditional Ising models.

Proof. Note that it suffices to prove the first equality, since the second equality may be obtained by averaging over Π_0. Further note that once we have proved the equality for uncensored infection vectors, the extension to censored vectors follows immediately from the additive expression for the censoring measure over all possible uncensored configurations.

We begin by focusing on the stochastic spreading model. The probability of an uncensored infection vector J is

$$\mathbb{P}_0(\mathcal{I} = J) = \sum_{P \in \mathfrak{P}(J)} \mathbb{P}_0(\mathcal{P} = P).$$

Thus, it suffices to prove that $\mathbb{P}_0(\mathcal{P} = P) = \mathbb{P}_0(\pi_0\mathcal{P} = P)$, for all $P \in \mathfrak{P}(J)$, where

$$\pi_0 P = \pi_0(P_1, \ldots, P_k) = (\pi_0(P_1), \ldots, \pi_0(P_k)).$$

The probability of a path P is equal to

$$\mathbb{P}_0(\mathcal{P} = P) = \prod_{t=1}^{k} \frac{1 + N_{t,\mathrm{in}}(P)}{n + 1 - t + \beta N_t(P)}.$$

Thus, it suffices to show that $N_{t,\text{in}}(P)$ and $N_t(P)$ are permutation-invariant. This is done in Lemma 8.2.2 below.

In the case of the Ising model, we have seen that the probability of an infection J only depends on its Hamiltonian. Lemma 8.2.3 below shows that the Hamiltonian statistic is permutation-invariant. □

We now provide the proofs of the supporting lemmas to Lemma 8.2.1. Note that the following lemmas are all deterministic statements in terms of a fixed graph $\mathcal{G}$, with the automorphism group $\Pi := \text{Aut}(\mathcal{G})$; they will be applied with $\mathcal{G} = \mathcal{G}_0$ and $\mathcal{G}_1$ in our arguments in the paper.

Lemma 8.2.2. *The statistics $N_{t,in}$ and N_t are Π-invariant when computed on graph $\mathcal{G}$.*

Proof. We have

$$N_{t,\text{in}}(P) = \sum_{s=1}^{t-1} \mathbf{1}\{(P_s, P_t) \in \mathcal{E}\},$$
$$N_t(P) = \sum_{(u,v)\in\mathcal{E}} \mathbf{1}\{\text{exactly one of } u \text{ or } v \text{ is in } (P_1, \ldots, P_{t-1})\}.$$

For $\pi \in \Pi$, we have

$$\begin{aligned} N_{t,\text{in}}(P) &= \sum_{s=1}^{t-1} \mathbf{1}\Big\{(P_s, P_t) \in \pi^{-1}\mathcal{E}\Big\} \\ &= \sum_{s=1}^{t-1} \mathbf{1}\{(\pi(P_s), \pi(P_t)) \in \mathcal{E}\} \\ &= N_{t,\text{in}}(\pi P), \end{aligned}$$

where we have used the definition of an automorphism and the closure of groups under inversion for the first equality. Similarly, we may write

$$\begin{aligned} N_t(P) &= \sum_{(u,v)\in\pi^{-1}\mathcal{E}} \mathbf{1}\{\text{exactly one of } u \text{ or } v \text{ is in } (P_1, \ldots, P_{t-1})\} \\ &= \sum_{(u,v)\in\mathcal{E}} \mathbf{1}\Big\{\text{exactly one of } \pi^{-1}(u) \text{ or } \pi^{-1}(v) \text{ is in } (P_1, \ldots, P_{t-1})\Big\} \\ &= \sum_{(u,v)\in\mathcal{E}} \mathbf{1}\{\text{exactly one of } u \text{ or } v \text{ is in } (\pi(P_1), \ldots, \pi(P_{t-1}))\} \\ &= N_t(\pi P). \end{aligned}$$

□

Lemma 8.2.3. *The edges-within statistic $W(J) = -\mathcal{H}(J)$ is Π-invariant when computed on graph $\mathcal{G}$.*

Proof. For $\pi \in \Pi$, we have

$$\begin{aligned} W(J) &= \sum_{(u,v)\in\mathcal{E}} \mathbf{1}\{J_u = J_v = 1\} \\ &= \sum_{(u,v)\in\pi^{-1}\mathcal{E}} \mathbf{1}\{J_u = J_v = 1\} \\ &= \sum_{(u,v)\in\mathcal{E}} \mathbf{1}\Big\{J_{\pi^{-1}(u)} = J_{\pi^{-1}(v)} = 1\Big\} \\ &= \sum_{(u,v)\in\mathcal{E}} \mathbf{1}\{(\pi J)_u = (\pi J)_v = 1\} \\ &= W(\pi J), \end{aligned}$$

using similar arguments to the proof of Lemma 8.2.2. □

8.2.3 Hunt-Stein

Proof of Theorem 7.4.3. Clearly, ψ^* is Π-invariant, since for any $\pi' \in \Pi$, we have

$$\psi^*(\pi'\mathcal{I}) = \frac{1}{|\Pi|}\sum_{\pi\in\Pi} \varphi^*(\pi\pi'\mathcal{I}) = \frac{1}{|\Pi|}\sum_{\pi\pi'\in\Pi\pi'} \varphi^*(\pi\pi'\mathcal{I}) = \frac{1}{|\Pi|}\sum_{\tilde{\pi}\in\Pi} \varphi^*(\tilde{\pi}\mathcal{I}) = \psi^*(\mathcal{I}).$$

We now show that ψ^* satisfies the inequality

$$\inf_{\pi\in\Pi} \mathbb{E}_{\pi\theta}[\varphi^*(\mathcal{I})] \le \mathbb{E}_\theta[\psi^*(\mathcal{I})] \le \sup_{\pi\in\Pi} \mathbb{E}_{\pi\theta}[\varphi^*(\mathcal{I})], \qquad \forall\theta \in \Theta. \tag{8.7}$$

We may write

$$\mathbb{E}_\theta[\psi^*(\mathcal{I})] = \mathbb{E}_\theta\left[\frac{1}{|\Pi|}\sum_{\pi\in\Pi} \varphi^*(\pi\mathcal{I})\right] = \frac{1}{|\Pi|}\sum_{\pi\in\Pi} \mathbb{E}_\theta[\varphi^*(\pi\mathcal{I})] = \frac{1}{|\Pi|}\sum_{\pi\in\Pi} \mathbb{E}_{\pi\theta}[\varphi^*(\mathcal{I})].$$

Furthermore, we clearly have

$$\inf_{\pi\in\Pi} \mathbb{E}_{\pi\theta}[\varphi^*(\mathcal{I})] \le \frac{1}{|\Pi|}\sum_{\pi\in\Pi} \mathbb{E}_{\pi\theta}[\varphi^*(\mathcal{I})] \le \sup_{\pi\in\Pi} \mathbb{E}_{\pi\theta}[\varphi^*(\mathcal{I})].$$

Combining the two relations gives inequality (8.7).

In particular, we have

$$\mathbb{E}_\theta[\psi^*(\mathcal{I})] \le \sup_{\pi\in\Pi} \mathbb{E}_{\pi\theta}[\varphi^*(\mathcal{I})] \le \alpha, \qquad \text{for all } \theta \text{ in } \Theta_0,$$

so ψ^* is a level-α test. Furthermore,

$$\mathbb{E}_\theta[\psi^*(\mathcal{I})] \ge \inf_{\pi\in\Pi} \mathbb{E}_{\pi\theta}[\varphi^*(\mathcal{I})], \qquad \text{for all } \theta \text{ in } \Theta_1,$$

so ψ^* is maximin. □

8.2.4 Likelihood Ratio

Proof of Theorem 7.4.1. We begin by considering the case $c = 0$. The likelihood under H_0 is

$$L(\mathcal{G}_{\text{empty}}, \beta; \mathcal{I}) = \sum_{P \in \mathfrak{P}(\mathcal{I})} \frac{1}{n} \cdots \frac{1}{n-(k-1)} = k! \cdot \frac{1}{n} \cdots \frac{1}{n-(k-1)}.$$

For $\mathcal{G}_1$, we have

$$\begin{aligned} L(\mathcal{G}_1, \beta; \mathcal{I}) &= \sum_{P \in \mathfrak{P}(\mathcal{I})} \prod_{t=1}^{k} \frac{1 + \beta N_{t,\text{in}}}{(n+1-t) + \beta N_t} \\ &= \sum_{P \in \mathfrak{P}(\mathcal{I})} \prod_{t=1}^{k} (1 + \beta N_{t,\text{in}}) \prod_{t=1}^{k} \frac{1}{(n+1-t) + \beta N_t}, \end{aligned} \tag{8.8}$$

where $N_{t,\text{in}}$ and N_t are computed with respect to $\mathcal{G}_1$. Thus, the likelihood ratio is

$$\begin{aligned} R(\mathcal{G}_{\text{empty}}, \mathcal{G}_1, \beta; \mathcal{I}) &= \frac{1}{k!} \sum_{P \in \mathfrak{P}(\mathcal{I})} \left(\prod_{t=1}^{k} (1 + \beta N_{t,\text{in}}) \right) \left(\prod_{t=1}^{k} \frac{n+1-t}{(n+1-t) + \beta N_t} \right) \\ &:= \frac{1}{k!} \sum_{P \in \mathfrak{P}(\mathcal{I})} A(P) B(P). \end{aligned}$$

We analyze each of the products $A(P)$ and $B(P)$ separately. We have the upper bound

$$\begin{aligned} A(P) &= \exp\left(\sum_{t=1}^{k} \log(1 + \beta N_{t,\text{in}}) \right) \\ &\leq \exp\left(\beta \sum_{t=1}^{k} N_{t,\text{in}} \right) \\ &= \exp(\beta W_1) \\ &= 1 + \beta W_1 + O(\beta^2 W_1^2), \end{aligned}$$

where we have used the fact that $\sum_{t=1}^{k} N_{t,\text{in}} = W_1$ when $N_{t,\text{in}}$ is computed along any possible infection path. Furthermore, we have the lower bound

$$\begin{aligned} A(P) &\geq \exp\left(\sum_{t=1}^{k} \beta N_{t,\text{in}} - \frac{1}{2} \beta^2 N_{t,\text{in}}^2 \right) \\ &= \exp\left(\beta W_1 - \frac{1}{2} \beta^2 \sum_{t=1}^{k} N_{t,\text{in}}^2 \right) \\ &\geq 1 + \beta W_1 - \frac{1}{2} \beta^2 \sum_{t=1}^{k} N_{t,\text{in}}^2 \\ &= 1 + \beta W_1 - O(\beta^2 W_1^2). \end{aligned}$$

Turning to $B(P)$, observe that we have the trivial upper bound $B(P) \le 1$. For a lower bound, we write

$$\begin{aligned} B(P) &= \exp\left(\sum_{t=1}^{k} \log(n+1-t) - \log(n+1-t+\beta N_t)\right) \\ &= \exp\left(-\sum_{t=1}^{k} \int_{n+1-t}^{n+1-t+\beta N_t} \frac{1}{x} dx\right) \\ &\ge \exp\left(-\sum_{t=1}^{k} \frac{\beta N_t}{n+1-t}\right) \\ &\ge 1 - \beta \sum_{t=1}^{k} \frac{N_t}{n+1-t} \\ &= 1 - \beta Q(P). \end{aligned}$$

Altogether, we have

$$\begin{aligned} R(\mathcal{G}_{\text{empty}}, \mathcal{G}_1, \beta; \mathcal{I}) &= \frac{1}{k!} \sum_{P \in \mathfrak{P}(\mathcal{I})} \Big(1 + \beta W_1 + O(\beta^2 W_1^2)\Big)\Big(1 - O(\beta Q(P))\Big) \\ &= \Big(1 + \beta W_1 + O(\beta^2 W_1^2)\Big)\left(1 - \frac{1}{k!} \sum_{P \in \mathfrak{P}(\mathcal{I})} O(\beta Q(P))\right), \end{aligned}$$

using the fact that $|\mathfrak{P}(\mathcal{I})| = k!$.

We now turn to the case when $c > 0$. Recall that

$$L(\mathcal{G}_{\text{empty}}, \beta; \mathcal{I}) = \frac{1}{\binom{n}{k\ c}}$$

and

$$L(\mathcal{G}_1, \beta; \mathcal{I}) = \frac{1}{\mu(\mathbb{I}_{k,c}; \beta)} \sum_{\mathcal{I}' \in \mathcal{U}(\mathcal{I})} \frac{1}{\binom{n}{c}} L^u(\mathcal{G}_1, \beta; \mathcal{I}'),$$

where L^u denotes the likelihood (8.8) computed without censoring. Let $\mathcal{U}(\mathcal{I}, c')$ denote the subset of $\mathcal{U}(\mathcal{I})$ where each infection has exactly $k + c'$ infected vertices (i.e, c' of the infected vertices were censored). Then $\mathcal{U}(\mathcal{I}) = \bigcup_{c'=0}^{c} \mathcal{U}(\mathcal{I}, c')$, and we may write

$$\begin{aligned} R(\mathcal{G}_{\text{empty}}, \mathcal{G}_1, \beta; \mathcal{I}) = \frac{\binom{n}{k\ c}}{\mu(\mathbb{I}_{k,c}; \beta)} \cdot \frac{1}{\binom{n}{c}} \sum_{c'=0}^{c} \sum_{\mathcal{I}' \in \mathcal{U}(\mathcal{I}, c')} \sum_{P \in \mathfrak{P}(\mathcal{I}')} \\ \left(\prod_{t=1}^{k+c'} (1 + \beta N_{t,\text{in}})\right)\left(\prod_{t=1}^{k+c'} \frac{n+1-t}{n+1-t+\beta N_t}\right) \frac{(n-k-c')!}{n!}. \end{aligned}$$

Again, denote the products inside the sums by $A(P)$ and $B(P)$, respectively. Since $\mathcal{I}'$ is an uncensored vector, the same argument as before gives

$$\begin{aligned} 1 + \beta W_1(\mathcal{I}') - O(\beta^2 W_1(\mathcal{I}')) \le A(P) \le 1 + \beta W_1(\mathcal{I}') + O\Big(\beta^2 W_1(\mathcal{I}')^2\Big), \\ 1 - \beta Q(P) \le B(P) \le 1. \end{aligned}$$

Hence, the likelihood ratio may be written as

$$\begin{aligned}
R(\mathcal{G}_{\text{empty}}, \mathcal{G}_1, \beta; \mathcal{I}) &= \frac{\binom{n}{k}\binom{n}{c}}{\mu(\mathbb{I}_{k,c}; \beta)} \cdot \frac{1}{\binom{n}{c}} \sum_{\mathcal{I}' \in \mathcal{U}(\mathcal{I})} \sum_{P \in \mathfrak{P}(\mathcal{I}')} \\
&\qquad \Big(1 + \beta W_1 + O(\beta^2 W_1(\mathcal{I}')^2)\Big)\Big(1 - O(\beta Q(P))\Big) \frac{(n - |\mathcal{I}'|)!}{n!} \\
&= \frac{\binom{n}{k}\binom{n}{c}}{\mu(\mathbb{I}_{k,c}; \beta)} \cdot \frac{1}{\binom{n}{c}} \sum_{\mathcal{I}' \in \mathcal{U}(\mathcal{I})} \\
&\qquad \frac{1}{\binom{n}{|\mathcal{I}'|}}\Big(1 + \beta W_1 + O(\beta^2 W_1(\mathcal{I}')^2)\Big)\left(1 - \frac{1}{|\mathcal{I}'|!} \sum_{P \in \mathfrak{P}(\mathcal{I}')} O(\beta Q(P))\right) \\
&= \sum_{\mathcal{I}' \in \mathcal{U}(\mathcal{I})} D(\beta, k, c, |\mathcal{I}'|)\Big(1 + \beta W_1(\mathcal{I}') + O(\beta^2 W_1(\mathcal{I}')^2)\Big) \\
&\qquad \times \left(1 - \frac{1}{|\mathcal{I}'|!} \sum_{P \in \mathfrak{P}(\mathcal{I}')} O(\beta Q(P))\right).
\end{aligned}$$

This completes the proof. □

Proof of Theorem 7.4.2. Let $\mathcal{I}$ and $\mathcal{I}'$ be two infections in $\mathbb{I}_{k,c}$. Then, there is some π in S_n such that $\mathcal{I} = \pi\mathcal{I}'$. Thus, we have

$$L(\theta; \mathcal{I}) = L(\theta; \pi\mathcal{I}') = L(\pi\pi^{-1}\theta; \pi\mathcal{I}') = L(\theta'; \mathcal{I}'),$$

where $\theta' = \pi^{-1}\theta$. Note that θ' is an element of Θ_0 by assumption. Thus, we conclude

$$\sup_{\theta \in \Theta_0} L(\theta, \mathcal{I}) = \sup_{\theta' \in \Theta_0} L(\theta', \mathcal{I}').$$

The rest of the proof follows immediately from the discussion preceding the theorem. □

8.2.5 Risk Bounds when $\mathcal{G}_0$ is the Star

Proof of Proposition 7.3.3. By Corollary 7.3.1, it suffices to compute the risk bound when the null hypothesis corresponds to the empty graph.

Let t_α denote the rejection threshold of the permutation test. Since t_α is defined to be the α-quantile of the edges-within statistic under the null hypothesis, we have

$$\mathbb{P}_0(W_1 > t_\alpha) \leq \alpha.$$

Thus, it remains to bound the Type II error. Our proof uses Lemma 8.5.1 to derive concentration of $W_1(\mathcal{I})$ to $\mathbb{E}[W_1(\mathcal{I})]$. Note that under both H_0 and H_1, we may apply Lemma 8.5.1 with X_i equal to the identity of the i^{th} uncensored infected node and $f(X_1, \ldots, X_k) = W_1(\mathcal{I})$.

Since each node is involved in at most D edges, we may take and $c_i = D$ for all $1 \le i \le k$. This leads to the following concentration bounds, which hold for all $t > 0$:

$$\begin{aligned} \mathbb{P}(W_1(\mathcal{I}) - \mathbb{E}[W_1(\mathcal{I})] \ge t) &\le \exp\left(-\frac{2t^2}{kD^2}\right), \\ \mathbb{P}(W_1(\mathcal{I}) - \mathbb{E}[W_1(\mathcal{I})] \le -t) &\le \exp\left(-\frac{2t^2}{kD^2}\right). \end{aligned} \tag{8.9}$$

We begin with the following lemma.

Lemma 8.2.4. *The rejection threshold satisfies the bound*

$$t_\alpha \le \frac{Dk(k-1)}{2(n-1)} + \sqrt{\frac{kD^2}{2}\log\frac{1}{\alpha}}.$$

Proof. We first compute $\mathbb{E}_0[W(\mathcal{I})]$. Let V_i denote the i^{th} uncensored vertex that is infected. We may write

$$\begin{aligned} \mathbb{E}_0[W(\mathcal{I})] &= \mathbb{E}\left[\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\mathbf{1}\{(V_i, V_j) \in \mathcal{E}_1\}\right] \\ &= \frac{k(k-1)}{2} \cdot \mathbb{E}[\mathbf{1}\{(V_1, V_2) \in \mathcal{E}_1\}] \\ &= \frac{k(k-1)}{2} \cdot \frac{|\mathcal{E}_1|}{\binom{n}{2}} \\ &= |\mathcal{E}_1|\frac{k(k-1)}{n(n-1)} \\ &= \frac{Dk(k-1)}{2(n-1)} \end{aligned}$$

Note that the last line uses the simple equality $|\mathcal{E}_1| = Dn/2$. Applying the bound (8.9) with

$$t = \sqrt{\frac{kD^2}{2}\log\frac{1}{\alpha}},$$

we then have

$$\mathbb{P}_0\left(W(\mathcal{I}) \ge \frac{Dk(k-1)}{2(n-1)} + \sqrt{\frac{kD^2}{2}\log\frac{1}{\alpha}}\right) \le \alpha,$$

implying the desired result. □

We now derive a lower bound for $\mathbb{E}_1[W(\mathcal{I})]$. We have the following result:

Lemma 8.2.5. *Let $\mathcal{G}_0$ be a vertex-transitive graph with degree D. Then we have the bound*

$$\mathbb{E}_1[W(\mathcal{I})] \ge \frac{D}{2} C_k H(\beta).$$

The proof of Lemma 8.2.5 is fairly technical and is contained in Section 8.4.1.

Combining the result of Lemma 8.2.5 with the concentration bound (8.9), we then have

$$\begin{aligned}\mathbb{P}_1(W_1(\mathcal{I}) \le t_\alpha) &\le \mathbb{P}_1\Big(W(\mathcal{I}) - \mathbb{E}[W(\mathcal{I})] \le t_\alpha - \frac{D}{2}C_k H(\beta)\Big) \\ &\le \exp\left\{-\frac{2}{kD^2}\left(\frac{D}{2}C_k H(\beta) - t_\alpha\right)^2\right\}. \end{aligned} \tag{8.10}$$

Finally, substituting the bound on t_α from Lemma 8.2.4 yields the required inequality. □

Proof of Proposition 8.1.1. Since the proof parallels the argument in Proposition 7.3.3, we only highlight the necessary modifications. In particular, inequalities (8.9) may be replaced by the following concentration bounds:

$$\begin{aligned}\mathbb{P}\Big(\overline{W} - \mathbb{E}[\overline{W}] \ge t\Big) &\le \exp\left(-\frac{2mt^2}{kD^2}\right), \\ \mathbb{P}\Big(\overline{W} - \mathbb{E}[\overline{W}] \le -t\Big) &\le \exp\left(-\frac{2mt^2}{kD^2}\right). \end{aligned} \tag{8.11}$$

This is due to the fact that we may apply Lemma 8.5.1 to the variables $\{X_{\ell,i}\}$, where $X_{\ell,i}$ denotes the identity of the i^{th} uncensored infected node in the ℓ^{th} spreading process, and $M_{Dm} = \overline{W} - \mathbb{E}[\overline{W}]$. We may take $c_{\ell,i} = D/m$ for all (ℓ, i).

The bound in Lemma 8.2.4 may then be replaced by the following bound on the rejection threshold:

$$t_\alpha \le \frac{Dk(k-1)}{2(n-1)} + \sqrt{\frac{kD^2}{2m}\log\frac{1}{\alpha}}.$$

Similarly, although Lemma 8.2.5 remains unchanged, the bound (8.10) will be modified with an additional factor of m appearing in the numerator of the exponent. □

8.2.6 Risk Bounds when $\mathcal{G}_1$ is the Star

Proof of Proposition 7.3.2. We first derive the maximum likelihood estimator. The likelihoods may be written as

$$L_i(\beta; \mathcal{I}) = \mathbb{P}_i(\mathcal{I}_1)\mathbb{P}_i(\mathcal{I}|\mathcal{I}_1).$$

Note that we have the equality

$$\mathbb{P}_0(\mathcal{I}|\mathcal{I}_1) = \mathbb{P}_1(\mathcal{I}|\mathcal{I}_1),$$

since under both hypotheses, given the infection status of vertex 1, all status assignments of the remaining nodes are equally likely. Hence, the MLE reduces to comparing $\mathbb{P}_0(\mathcal{I}_1)$ and $\mathbb{P}_1(\mathcal{I}_1)$.

We have

$$\mathbb{P}_0(\mathcal{I}_1 = 1) = \frac{k}{n}, \quad \mathbb{P}_0(\mathcal{I}_1 = 0) = \frac{n-k}{n},$$

whereas

$$\mathbb{P}_1(\mathcal{I}_1 = 1) > \frac{k}{n}, \quad \mathbb{P}_1(\mathcal{I}_1 = 0) < \frac{n-k}{n},$$

since the center of the star is more likely to be infected relative to the leaves. Hence, the test that rejects H_0 according to the center indicator statistic $\mathbf{1}\{\mathcal{I}_1 = 1\}$ is indeed a maximum likelihood estimator. Note that when $\mathcal{I}_1 = \star$, we may make an arbitrary decision, so we decide to default to H_0 in that case. Finally, observe that the Type I error is controlled by α when $\alpha \geq k/n$. □

Proof of Proposition 7.3.3. We begin with the following lemma, proved in Section 8.4.2.

Lemma 8.2.6. *Under the hypothesis that the graph $\mathcal{G}_1$ is a star, we have the bounds*

$$\mathbb{P}_1(\mathcal{I}_1 = 0) \geq \exp\left(-\frac{k + \beta k(k-1)/2}{n-k}\right),$$

and

$$\mathbb{P}_1(\mathcal{I}_1 = 0) \leq \begin{cases} \exp\left(-\frac{k+\beta k(k-1)/2}{(n-k+1)+(k-1)\beta}\right), & \text{if } \beta \geq 1, \\ \exp\left(-\frac{k+\beta k(k-1)/2}{n}\right), & \text{if } \beta < 1. \end{cases}$$

Returning to the proof of the proposition, note that

$$R_{k,0}(C, \beta) = \mathbb{P}_0(\mathcal{I}_1 = 1) + \mathbb{P}_1(\mathcal{I}_1 = 0) = \frac{k}{n} + \mathbb{P}_1(\mathcal{I}_1 = 0).$$

Applying the bounds in Lemma 8.2.6 then implies the desired result. □

Proof of Proposition 8.1.2. By the analog of Corollary 7.3.1 for multiple spreading processes, it suffices to consider the risk when $\mathcal{G}_0$ is the empty graph. Let t_α be the level α threshold. We wish to bound

$$R_{k,0}(P_{\overline{C},\alpha}, \beta) = \mathbb{P}_0\Big(\overline{C} > t_\alpha\Big) + \mathbb{P}_1\Big(\overline{C} \leq t_\alpha\Big) \leq \alpha + \mathbb{P}_1\Big(\overline{C} \leq t_\alpha\Big).$$

To bound the Type II error, it suffices to pick any threshold t'_α such that

$$\mathbb{P}_0\Big(\overline{C} > t'_\alpha\Big) \leq \alpha. \tag{8.12}$$

By definition, this guarantees that $t_\alpha \leq t'_\alpha$, and as a consequence,

$$\mathbb{P}_1\Big(\overline{C} \leq t_\alpha\Big) \leq \mathbb{P}_1\Big(\overline{C} \leq t'_\alpha\Big).$$

Accordingly, let

$$t'_\alpha = \frac{k}{n} + \sqrt{\frac{1}{2m} \log \frac{1}{\alpha}}.$$

By applying Hoeffding's inequality, we see that t'_α satisfies inequality (8.12).

We will apply Hoeffding's inequality again to bound $\mathbb{P}_1(\overline{C} < t'_\alpha)$. Note that

$$\mathbb{E}_1[\overline{C}] = \mathbb{P}_1(\mathcal{I}_1 = 1) \geq p_{k,0}(\beta),$$

by Lemma 8.2.6. It follows that

$$\begin{aligned}\mathbb{P}_1(\overline{C} \leq t'_\alpha) &\leq \mathbb{P}_1\Big(\overline{C} - \mathbb{E}_1[\overline{C}] \leq t_\alpha - p_{k,0}(\beta)\Big) \\ &\leq \exp\left(-2m\left(\frac{k}{n} + \sqrt{\frac{1}{2m}\log\frac{1}{\alpha}} - p_{k,0}(\beta)\right)^2\right),\end{aligned}$$

implying the desired result. □

8.3 Proofs of Corollaries

In this section, we provide proofs of corollaries stated in Chapter 7 and Chapter 8.

8.3.1 Proof of Corollary 7.3.1

Let R_0 and R'_0 denote the risks under null hypotheses H_0 and H'_0, respectively, and let $\mathcal{A}$ denote the rejection region of the test statistic. We have

$$\begin{aligned}R_0 &= \mathbb{P}_0(S(\mathcal{I}) \in \mathcal{A}) + \mathbb{P}_1(S(\mathcal{I}) \notin \mathcal{A}), \quad \text{and} \\ R'_0 &= \mathbb{P}'_0(S(\mathcal{I}) \in \mathcal{A}) + \mathbb{P}_1(S(\mathcal{I}) \notin \mathcal{A}),\end{aligned}$$

where $\mathbb{P}'_0$ denotes the probability distribution under H'_0.

Note that $\Pi_1\Pi_0 = S_n$ by assumption, and also $\Pi_1\Pi'_0 = S_n$, since $\Pi'_0 = S_n$. By Theorem 7.2.1, we then have

$$\mathbb{P}_0(S(\mathcal{I}) \in \mathcal{A}) = \frac{1}{n!}\sum_{\pi \in S_n} \mathbb{P}(S(\pi J) \in \mathcal{A}) = \mathbb{P}'_0(S(\mathcal{I}) \in \mathcal{A}),$$

for any fixed infection vector $J \in \mathbb{I}_{k,c}$. It follows that $R_0 = R'_0$, as claimed.

8.3.2 Proof of Corollary 7.3.2

Suppose $\mathcal{G}_0$ is the star and $\mathcal{G}_1$ is vertex-transitive. Then Π_1 contains permutations g_i mapping vertex 1 to vertex i, for $1 \leq i \leq n$. Let $G = \{g_1, \ldots, g_n\}$, and note that $G \cap \Pi_0 = \{g_1\}$. Furthermore, the cosets $g_i\Pi_0$ are unique. Finally, by equation (8.1), we have

$$|G\Pi_0| = \frac{|G||\Pi_0|}{|G \cap \Pi_0|} = \frac{n(n-1)!}{1} = n!.$$

Thus, we conclude that $\Pi = S_n$. The proof when $\mathcal{G}_1$ is the star graph is analogous.

8.3.3 Proof of Corollary 7.3.3

We first show that $|\mathcal{C}_k(u,v)| = (k-1)2^{k-1}$, for any edge (u,v). Note that the number of possible choices for the k infected vertices in a cascade involving u and v is $k-1$, corresponding to segments of k neighboring nodes in the cycle graph. Furthermore, the number of orderings of infected vertices in the segment is 2^{k-1}, corresponding to whether the infection proceeds to the right or left on each step.

Substituting $C_k = (k-1)2^{k-1}$ and

$$H(\beta) = \prod_{m=1}^{k-1} \frac{\beta}{n-m+2\beta}$$

into the risk bound in Proposition 7.3.1 yields the first part of the corollary.

For the second and third part of the corollary, note that

$$\lim_{\beta\to\infty} H(\beta) = \prod_{m=1}^{k-1} \frac{1}{2} = 2^{-(k-1)}.$$

Simple algebra then yields the desired results.

8.3.4 Proof of Corollary 8.1.2

The proof of this corollary refers back to the proof of Proposition 7.3.1. Let $|\mathcal{E}_1'| = n$ denote the number of edges in the cycle graph $\mathcal{G}_1'$. If $D' = 2$ denotes the maximum degree of $\mathcal{G}_1'$, we still have the bound

$$t_\alpha \le |\mathcal{E}_1'| \frac{k(k-1)}{n(n-1)} + \sqrt{\frac{k(D')^2}{2} \log \frac{1}{\alpha}},$$

from Lemma 8.2.4. The analog of Lemma 8.2.5, specialized to the case of an infection spreading over the path graph, is the following bound:

Lemma 8.3.1. *Under the alternative hypothesis that $\mathcal{G}_1$ is the path graph, we have*

$$\mathbb{E}_1[W_{\mathcal{G}_1'}(\mathcal{I})] \ge \frac{(n-c)(n-c-1)}{n^2(n-1)} \cdot (k-1)2^{k-1} \cdot \frac{n-k+1}{n} \cdot \prod_{m=1}^{k-1} \frac{\beta}{n-m+2\beta}.$$

The proof of Lemma 8.3.1 is provided in Section 8.4.3.

Finally, we have the concentration inequalities

$$\mathbb{P}_0\Big(W_{\mathcal{G}_1'}(\mathcal{I}) - \mathbb{E}_0\Big[W_{\mathcal{G}_1'}(\mathcal{I})\Big] \ge t\Big) \le \exp\left(-\frac{2t^2}{k(D')^2}\right),$$
$$\mathbb{P}_1\Big(W_{\mathcal{G}_1'}(\mathcal{I}) - \mathbb{E}_1\Big[W_{\mathcal{G}_1'}(\mathcal{I})\Big] \le -t\Big) \le \exp\left(-\frac{2t^2}{k(D')^2}\right).$$

Combining the pieces as in the proof of Proposition 7.3.1 then yields the desired bound.

8.4 Proofs of Supporting Lemmas

Finally, we provide proofs of supporting technical lemmas.

8.4.1 Proof of Lemma 8.2.5

We begin by writing

$$\begin{aligned}\mathbb{E}_1[W(\mathcal{I})] &= \sum_{(u,v)\in\mathcal{E}_1} \mathbb{P}_1(\mathcal{I}_u = \mathcal{I}_v = 1) \\ &\geq \sum_{(u,v)\in\mathcal{E}_1} |\mathcal{C}_k(u,v)| \cdot \frac{\binom{n-2}{c}}{\binom{n}{c}} \cdot \frac{1}{n} \prod_{m=1}^{k-1} \frac{\beta}{n-m+\beta(2+m(D-2))}.\end{aligned}$$

Indeed, the inequality comes from restricting our consideration to infections where the $k-1$ nodes after the first are infected along edges of the graph rather than by random infection, and u and v are in the initial infection set. The term $|\mathcal{C}_k(u,v)|$ counts the number of such infection paths, and the term $\binom{n-2}{c}/\binom{n}{c}$ computes the probability that u and v will remain uncensored at the end of the process. Finally, the factor

$$\prod_{m=1}^{k-1} \frac{\beta}{n-m+\beta(2+m(D-2))}$$

lower-bounds the probability that each of the $k-1$ vertices after the first contracts the disease from one of its infected neighbors. Note that at each stage, the total number of edges connecting the $n-m$ uninfected nodes to the m previously infected nodes is bounded above by $2+m(D-2)$, since the infected nodes are necessarily connected to each other.

Recalling the definition of $H(\beta)$, we then have

$$\begin{aligned}\mathbb{E}_1[W(\mathcal{I})] &\geq \frac{(n-c)(n-c-1)}{n^2(n-1)} \cdot H(\beta) \cdot \sum_{(u,v)\in\mathcal{E}_1} |\mathcal{C}_k(u,v)| \\ &\geq |\mathcal{E}_1| \frac{(n-c)(n-c-1)}{n^2(n-1)} C_k H(\beta),\end{aligned}$$

as wanted.

8.4.2 Proof of Lemma 8.2.6

Clearly, we have

$$\mathbb{P}_1(\mathcal{I}_1 = 0) = \mathbb{P}_1(\mathcal{I}_1 \neq \star)\mathbb{P}_1(\mathcal{I}_1 = 0 | \mathcal{I}_1 \neq \star) = \left(\frac{n-c}{n}\right)\mathbb{P}_1(\mathcal{I}_1 = 0 | \mathcal{I}_1 \neq \star).$$

The latter probability is easier to calculate, since we may consider a process where we first choose c of the vertices $\{2,\ldots,n\}$ to censor, and then compute the probability that the k infected nodes lying in the remaining vertex set are all leaf nodes. Since vertex 1 is not infected, the spreading process is agnostic to the infection status of the c censored nodes.

We first consider the lower bound. We have

$$\begin{aligned}
\mathbb{P}_1\{\mathcal{I}_1 = 0 | \mathcal{I}_1 \neq \star\} &= \prod_{j=0}^{k-1} \frac{n-c-1-j}{(n-c-j)+j\beta} \\
&= \exp\left(\sum_{j=0}^{k-1} \log \frac{n-c-1-j}{(n-c-j)+j\beta} \right) \\
&= \exp\left(-\sum_{j=0}^{k-1} \int_{n-c-j-1}^{(n-c-j)+j\beta} \frac{1}{x} dx \right) \\
&\geq \exp\left(-\sum_{j=0}^{k-1} \frac{1+j\beta}{n-c-j-1} \right) \\
&\geq \exp\Big(-\frac{k+\beta k(k-1)/2}{n-c-k} \Big).
\end{aligned}$$

The upper bound may be derived in an analogous fashion. We have

$$\begin{aligned}
\mathbb{P}_1\{\mathcal{I}_1 = 0 | \mathcal{I}_1 \neq \star\} &= \exp\left(-\sum_{j=0}^{k-1} \int_{n-c-j-1}^{(n-c-j)+j\beta} \frac{1}{x} dx \right) \\
&\leq \exp\left(-\sum_{j=0}^{k-1} \frac{1+j\beta}{(n-c)+j(\beta-1)} \right).
\end{aligned}$$

When $\beta \geq 1$, the denominator is maximized for $j = k-1$; when $\beta < 1$, the denominator is maximized for $j = 0$. In the first case, we have

$$\begin{aligned}
\mathbb{P}_1\{\mathcal{I}_1 = 0 | \mathcal{I}_1 \neq \star\} &\leq \exp\left(-\frac{1}{(n-c)+(k-1)(\beta-1)} \sum_{j=0}^{k-1} (1+j\beta) \right) \\
&= \exp\Big(-\frac{k+\beta k(k-1)/2}{(n-c-k+1)+(k-1)\beta} \Big).
\end{aligned}$$

In the second case, we have

$$\mathbb{P}_1\{\mathcal{I}_1 = 0 | \mathcal{I}_1 \neq \star\} \leq \exp\left(-\frac{1}{n-c} \sum_{j=0}^{k-1} (1+j\beta) \right) = \exp\Big(-\frac{k+\beta k(k-1)/2}{n-c} \Big).$$

8.4.3 Proof of Lemma 8.3.1

As in the proof of Lemma 8.2.5, we begin by writing

$$\begin{aligned}
\mathbb{E}_1\Big[W_{\mathcal{G}_1'}(\mathcal{I}) \Big] &= \sum_{(u,v) \in \mathcal{E}_1'} \mathbb{P}_1(\mathcal{I}_u = \mathcal{I}_v = 1) \\
&\geq \sum_{(u,v) \in \mathcal{E}_1'} |\mathcal{C}_k'(u,v)| \cdot \frac{\binom{n-2}{c}}{\binom{n}{c}} \cdot \frac{1}{n} \prod_{m=1}^{k-1} \frac{\beta}{n-m+2\beta},
\end{aligned}$$

where $\mathcal{C}'_k(u,v)$ denotes the set of cascades involving (u,v) in $\mathcal{G}'_1$.

We claim that

$$\sum_{(u,v)\in\mathcal{E}'_1} |\mathcal{C}'_k(u,v)| = (k-1)(n-k+1)\cdot 2^{k-1}, \tag{8.13}$$

from which the result follows. Indeed, for $(u,v) = (i,i+1)$, with $1 \le i \le k-1$, we have

$$|\mathcal{C}'_k(u,v)| = i \cdot 2^{k-1},$$

since we have i choices for the collection of infected vertices in the cascade, and given a collection of vertices, the infection may spread according to 2^{k-1} different orderings. Similarly, we may argue that

$$\begin{aligned} |\mathcal{C}'_k(u,v)| &= (n-i)\cdot 2^{k-1}, \qquad \text{for } n-k+1 \le i \le n-1, \\ |\mathcal{C}'_k(u,v)| &= (k-1)\cdot 2^{k-1}, \qquad \text{for } k \le i \le n-k. \end{aligned}$$

Summing up over all choices of (u,v) then yields

$$\begin{aligned} \sum_{(u,v)\in\mathcal{G}'_1} |\mathcal{C}'_k(u,v)| &= k(k-1)\cdot 2^{k-1} + (k-1)(n-2k+1)\cdot 2^{k-1} \\ &= (k-1)(n-k+1)\cdot 2^{k-1}, \end{aligned}$$

which is equation (8.13).

8.5 Auxiliary Lemmas

For our risk bounds, we need a standard concentration result:

Lemma 8.5.1. *Suppose $\{M_t\}_{t=1}^T$ is a martingale with respect to some filtration and the differences $M_t - M_{t-1}$ have expectation 0 and are bounded by c_t. Then*

$$\mathbb{P}(M_T \ge t) \le \exp\left(-\frac{2t^2}{\sum_{t=1}^T c_t^2}\right),$$

and

$$\mathbb{P}(M_T \le -t) \le \exp\left(-\frac{2t^2}{\sum_{t=1}^n c_t^2}\right).$$

We will apply Lemma 8.5.1 to the statistic W_1 in the following manner: Let $\mathcal{P}_{1:t}$ denote the the first t infected vertices, and define $W_1(\mathcal{P}_{1:t})$ to be the edges-within statistic for the infection J in $\mathbb{I}_{t,0}$ corresponding to the partial path. Let $\mathcal{F}_t = \sigma(\mathcal{P}_{1:t})$ be the sigma-field of infections up to time t. Finally, define the martingale

$$M_t = W_1(\mathcal{P}_{1:t}) - \mathbb{E}[W_1(\mathcal{P}_{1:t})] = \sum_{s=1}^{t} (W_1(\mathcal{P}_{1:s}) - W_1(\mathcal{P}_{1:s-1})) - \mathbb{E}[W_1(\mathcal{P}_{1:s}) - W_1(\mathcal{P}_{1:s-1})],$$

where $W_1(\mathcal{P}_{1:0}) = 0$. Since $W_1(\mathcal{P}_{1:s}) - W_1(\mathcal{P}_{1:s-1})$ is in $[0,D]$ where D is the maximum degree of the graph, the martingale differences satisfy

$$|\Delta_t| = |M_t - M_{t-1}| = \Big|(W_1(\mathcal{P}_{1:t}) - W_1(\mathcal{P}_{1:t-1})) - \mathbb{E}[W_1(\mathcal{P}_{1:t}) - W_1(\mathcal{P}_{1:t-1})]\Big| \le D.$$

Thus, Lemma 8.5.1 applies with $c_t = D$.

9. Numerical Details

The numerical results are contained in this chapter. We consider an analysis of the HIV graph data and a battery of simulation results.

9.1 HIV Graph Analysis

We now detail our HIV graph analysis. We first applied Algorithm 7.2.2 with the edges-within statistic W, resulting in $W(\mathcal{I}) = 98$. Based on $B = 1000$ randomly drawn permutations, we obtained an approximate p-value of 0; i.e. none of the simulations yielded a value of W of at least 98. This should provide substantial evidence of the explanatory power of the graph. However, a closer examination of the network reveals that the censored nodes might not be chosen at random. To further validate this observation, we simulated our stochastic spreading model on the HIV graph, and computed W for $\lambda = 1$ and various values of β in $[10, 1000]$. A boxplot is provided in Figure 9.1. In particular, none of the simulations produced a value of W greater than or equal to 98, and increasing β even further seems to have little effect.

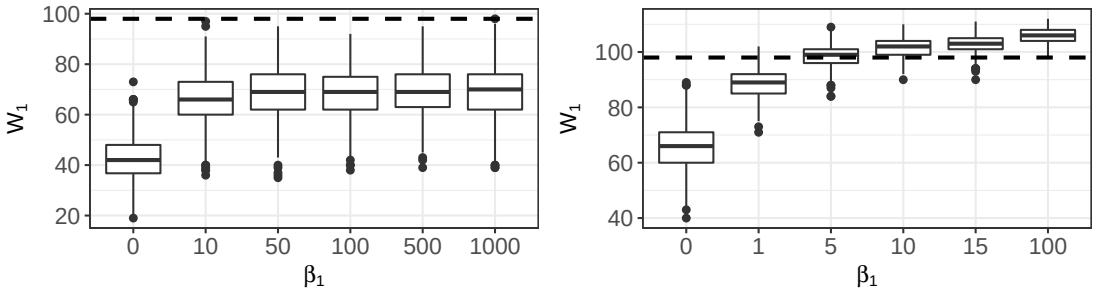

Figure 9.1 The distributions of W for different values of β on the HIV graph, based on 1000 simulations. The left plot corresponds to random censoring, whereas the right plot conditions on the censored nodes. The dashed line shows $W_1 = 98$, the value from the data. No β appears to be consistent with the observed value $W = 98$ in the randomized censoring model; such a value is more typical in the conditional censoring model.

Subsequently, we used Algorithm 8.1.2 with statistic W, with the empty graph as the null. The approximate p-value computed from 1000 simulations was again 0. However, simulating the stochastic spreading model on the HIV graph, conditioned on the censored nodes, leads to more reasonable values of W. Boxplots for simulations corresponding to $\lambda = 1$ and $\beta \in \{0, 1, 5, 10, 15, 100\}$ are shown in the right-hand plot of Figure 9.1. Note that for moderate values of β, the observed value 98 lies squarely within the range of the empirical distribution of W. This suggests that the model which conditions on the censored nodes provides a better fit to the data.

Finally, note that neither of the algorithms from Milling et al. (2015) produce meaningful

conclusions. In the case of the TB algorithm, all positive integers d lead to a threshold larger than the diameter of the graph, so the algorithm will always reject the null hypothesis. Similarly, the threshold for the TT algorithm is 779, which exceeds the weight of any spanning tree since the graph has only 250 vertices. The TB and TT algorithms both fail due to the closely-connected structure of HIV graph, which is a characteristic of many real-world networks, and the inability to select a good threshold.

9.2 Simulations

Simulations were conducted on various graphs to illustrate different phenomena. The first set of simulations concerns testing vertex-transitive graphs against the star graph to validate our theory. The second set of simulations corresponds to irregular graphs obtained from online social networks, where we use the empty graph as the null graph. The third set of simulations covers two cases of randomly-generated graphs. In the first case, our theory predicts that permutation test succeeds on average, whereas the second case corresponds to a departure from the conditions required for our theory to succeed, and shows that the permutation test might fail to control the Type I error as edge correlations increase.

We performed simulations to assess the performance of our permutation test. Our main interest was to explore the power of our test on various graphs. We compared the algorithms proposed in Milling et al. (2015), computed with respect to the following test statistics:

1. The *infection radius* $R(\mathcal{I})$, defined by

$$R_{\mathcal{G}}(J) := \min_{v \in \mathcal{V}} \left\{ \max_{u \in J^1} d_{\mathcal{G}}(u, v) \right\},$$

   where $d_{\mathcal{G}}(u, v)$ is the length of the shortest path connecting u to v in $\mathcal{G}$.

2. The *Steiner minimal tree*, defined for a subset of vertices M in a connected graph $\mathcal{G}$ to be a subtree of minimal edge weight containing all vertices in M (Gilbert and Pollak, 1968). Here, M is the set of infected vertices and each edge has unit weight.

   We are interested in a hypothesis test computed with respect to the edge weight of the Steiner minimal tree. However, finding the Steiner minimal tree is NP-complete (Garey and Johnson, 1979), so the statistic used for simulations is based on an approximate Steiner minimal tree, computed via a tractable algorithm due to Mehlhorn (1988). Let $T(\mathcal{I})$ be the edge weight of the approximate Steiner minimal tree.

The algorithms of Milling et al. (2015) reject the null hypothesis if the statistic $R(\mathcal{I})$ or $T(\mathcal{I})$, evaluated on $\mathcal{G}_1$, exceeds a certain threshold. The corresponding tests are called the Threshold Ball (TB) and Threshold Tree (TT) algorithms. The threshold for the TB algorithm is computed theoretically to be $1.1d^2(kn\log(\log n)/(n-c))^{\frac{1}{d}}$, when $\mathcal{G}_1$ is a toroidal grid of dimension d. For non-toroidal graphs, no threshold is provided, so our approach was to try to select a value of d appropriate to the threshold calculation. In the case of the TT algorithm, the threshold suggested by Milling et al. (2015) is $kn(\log\log n)^3/(n-c)$.

Note that both R and T are defined in relation to the topology of $\mathcal{G}_1$, so they are Π_1-invariant. Thus, we may apply Algorithm 7.2.2 and compare the results of the permutation test directly to the thresholding algorithms proposed by Milling et al. (2015). We write $\mathcal{A}_{\text{perm}}(S, B, \alpha)$ to denote the result of applying Algorithm 7.2.2 with Π_1-invariant statistic S, based on B randomly drawn permutations, with Type I error level α. We write $\mathcal{A}_{TB}(d)$ and $\mathcal{A}_{TT}$ to denote the TB algorithm with parameter d and the TT algorithms, respectively.

9.2.1 Vertex-Transitive Graphs

For the first set of simulations, we let $\mathcal{G}_0$ be the empty graph and $\mathcal{G}_1$ the 2-dimensional toroidal grid on n vertices. We set $n \in \{2500, 10000\}$ and $k = c = 500$. The results are based on 1000 simulations, for $\lambda = 1$ and $\beta \in \{1, 10, 100\}$. Table 9.1 contains the numerical results.

Algorithm 7.2.2 with the edges-within statistic W performed significantly better than all other algorithms in terms of Type II error. Note that the Type II error for Algorithm 7.2.2 with statistics R and T decreases as β increases. On the other hand, the TB and TT algorithms are uninformative for both grid sizes, since the thresholds set by the algorithms result in a rejection rule that never rejects the null hypothesis. This emphasizes an important feature of our permutation test, which always provides Type I error control, in contrast to the methods of Milling et al. (2015), which only guarantee that the Type I error will converge to 0 as n grows. Further note that our simulations involve a parameter $\lambda > 0$, which is not covered by the theory of Milling et al. (2015). On the other hand, for large β, most infected nodes should contract the disease from a neighbor rather than exogenous sources.

9.2.2 Online Social Network

In order to gauge the performance of our algorithm on more "irregular" networks, we performed simulations on real social network graphs with the empty graph as the null. We used two connected components of the Facebook social network from the Stanford Network Analysis Project (Leskovec and Krevl, 2014). The graphs, which we refer to as Facebook 1 and Facebook 2, correspond to users who downloaded the Social Circles app and can be seen in Figure 9.2.

We simulated 1000 infections on the graphs with $k = 200$, $c = 300$, $\lambda = 1$, and $\beta \in \{1, 10, 100\}$. For the TB algorithm, we chose the integer d providing the smallest threshold. Table 9.2 details the results. As seen in the table, Algorithm 7.2.2 with statistic W again performed better than the other algorithms in terms of Type II error. The TB and TT algorithms performed poorly, since the thresholds for all values of d are too large.

9.2.3 Erdős-Renyi versus Stochastic Block Model Graphs

For our first set of random graph simulations, we considered an infection spreading on an Erdős-Renyi graph, versus a two-block stochastic block model with equal-sized partitions,

Size	Algorithm	Threshold	Type I Error	Type II Error by β		
				1	10	100
$n = 2,500$	$\mathcal{A}_{\text{perm}}(W,\ 100,\ 0.01)$	220	0.028	0.000	0.000	0.000
	$\mathcal{A}_{\text{perm}}(R,\ 100,\ 0.01)$	45	0.150	0.517	0.000	0.000
	$\mathcal{A}_{\text{perm}}(T,\ 100,\ 0.01)$	1186	0.009	0.858	0.050	0.000
	$\mathcal{A}_{TB}(2)$	157	1.000	0.000	0.000	0.000
	$\mathcal{A}_{TT}$	5441	1.000	0.000	0.000	0.000
$n = 10,000$	$\mathcal{A}_{\text{perm}}(W,\ 100,\ 0.01)$	63	0.026	0.000	0.000	0.000
	$\mathcal{A}_{\text{perm}}(R,\ 100,\ 0.01)$	87	0.001	1.000	0.915	0.000
	$\mathcal{A}_{\text{perm}}(T,\ 100,\ 0.01)$	2,509	0.008	0.970	0.206	0.000
	$\mathcal{A}_{TB}(2)$	150	1.000	0.000	0.000	0.000
	$\mathcal{A}_{TT}$	5,760	1.000	0.000	0.000	0.000

Table 9.1 Threshold, Type I error, and Type II error for algorithms on the two-dimensional toroidal grid graph of size n. Here, $\mathcal{G}_0$ is the empty graph. We set $k = c = 500$. The statistic W performs best by far, and the permutation tests offer more reasonable error bounds. The Type I and Type II errors are approximated using 1000 simulations each.

Network	Algorithm	Threshold	Type I Error	Type II Error by β		
				1	10	100
Facebook 1	$\mathcal{A}_{\text{perm}}(W,\ 200,\ 0.01)$	46	0.005	0.242	0.000	0.000
	$\mathcal{A}_{\text{perm}}(R,\ 200,\ 0.01)$	11	0.018	0.958	0.798	0.307)
	$\mathcal{A}_{\text{perm}}(T,\ 200,\ 0.01)$	273	0.004	0.700	0.000	0.000
	$\mathcal{A}_{TB}(3)$	77	1.000	0.000	0.000	0.000
	$\mathcal{A}_{TT}$	1964	1.000	0.000	0.000	0.000
Facebook 2	$\mathcal{A}_{\text{perm}}(W,\ 100,\ 0.01)$	51	0.017	0.048	0.000	0.000
	$\mathcal{A}_{\text{perm}}(R,\ 100,\ 0.01)$	10	0.007	0.987	0.830	0.339
	$\mathcal{A}_{\text{perm}}(T,\ 100,\ 0.01)$	249	0.029	0.290	0.000	0.000
	$\mathcal{A}_{TB}(3)$	78	1.000	0.000	0.000	0.000
	$\mathcal{A}_{TT}$	1965	1.000	0.000	0.000	0.000

Table 9.2 Threshold, Type I error, and Type II error for algorithms when $\mathcal{G}_1$ is one of the subsets of the Facebook graph. Here, we set $k = 200$ and $c = 300$. Note that $\mathcal{A}_{\text{perm}}(W, B, \alpha)$ seems to perform better than the other algorithms.

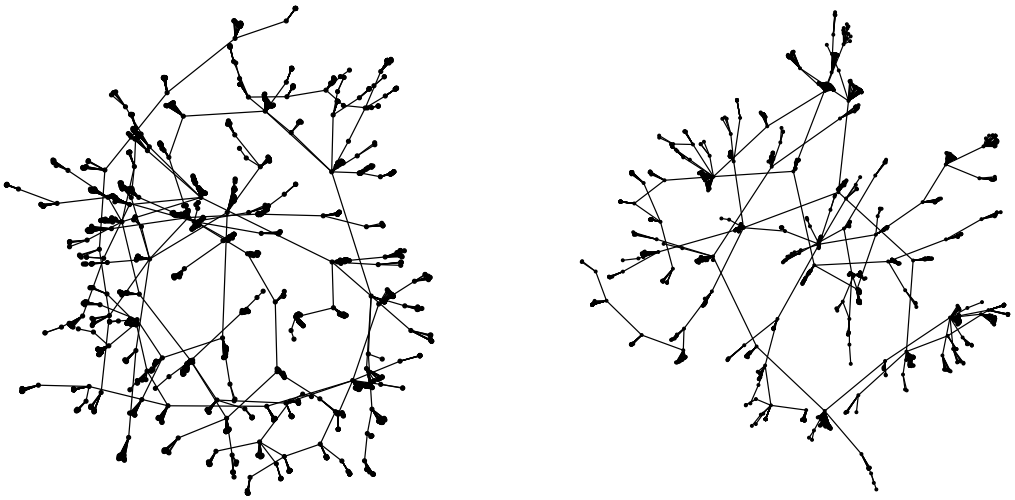

Figure 9.2 The Facebook 1 and Facebook 2 graphs, with 2040 and 1567 nodes. Note that the graphs contain a large proportion of leaves.

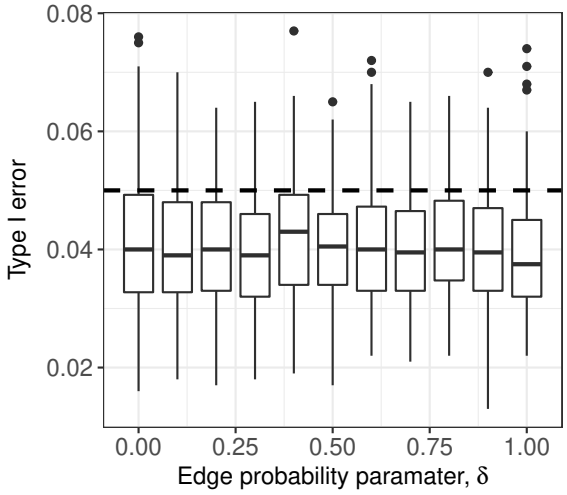

Figure 9.3 The Type I error of the permutation test applied to the test of an Erdős-Renyi random graph versus a two-block stochastic block model graph on $n = 500$ vertices and $k = c = 50$. The edge density is $p = \log(n)/n$, and the within-block connectivity is $a = 2\delta p$ and $\beta = 1$. On average, the permutation test controls the Type I error at the nominal level $\alpha = 0.05$.

where $n = 500$. We let $k = c = 50$. In the null graph, edges exist independently with probability $p = \log(n)/n$. In the alternative graph, the probability that an edge exists between nodes in the same partition is a, and the probability that an edge exists between nodes in different partitions is b, where $a/2 + b/2 = p$. We let $a = 2\delta p$ and varied δ in $[0, 1]$ in increments of 0.1. To estimate thresholds and error probabilities, we sampled each infection process 1000 times. We drew 100 pairs of random graphs for each δ.

Lemma 8.1.1 predicts the validity of the permutation test in controlling Type I error.

Furthermore, we expect the Type II error to decrease as the infections become sufficiently dissimilar through increasing β or δ. The first of these conclusions is verified by examining Figure 9.3: As δ varies, the permutation test controls the Type I error at a level $\alpha = 0.05$ on average, and with reasonably high probability. Furthermore, even outliers had Type I errors of no more than 0.08.

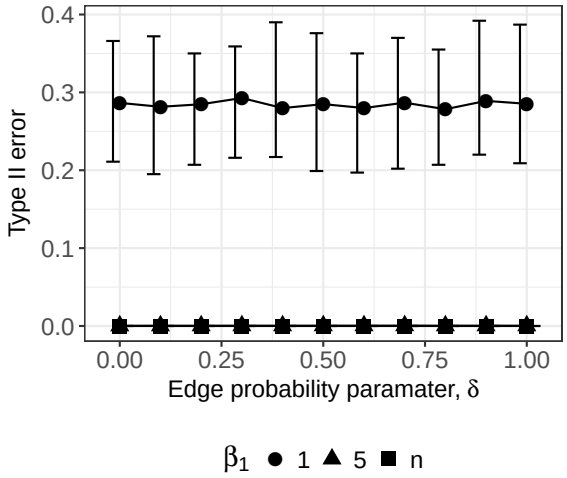

Figure 9.4 The Type II error of the permutation test applied to the test of an Erdős-Renyi random graph versus a two-block stochastic block model graph on $n = 500$ vertices and $k = c = 50$. The edge density is $p = \log(n)/n$, and the within-block connectivity is $a = 2\delta p$. The markers denote the average error, and the bars denote the range over the 100 simulated graphs. Perhaps surprisingly, the Type II error does not seem to depend on δ. However, the Type II error decreases with β_1.

To examine the Type II error, we provide two plots. First, Figure 9.4 shows the average Type II error for various levels of β_1 and δ, when $\beta_0 = 0$. The marker indicates the average value across the 100 graph realizations, and the bars show the range of Type II errors. Additionally, the bars near each point indicate the range of the Type II errors across simulated graphs. Note that the Type II error is low for even modest values of β_1. The second plot is Figure 9.5. Here, the mean Type II error is plotted for various levels of β_1 and δ, where $\beta_0 = n$. Additionally, the bars near each point indicate the range of Type II errors over simulated graphs. Note that Figures 9.4 and 9.5 are very similar. Thus, even strong spreading on the Erdős-Renyi random graph does not seem to affect Type II error both in terms of the average and the range of values.

9.2.4 Correlated Erdős-Renyi Random Graphs

We now consider a case where the permutation test performs poorly, by examining correlated Erdős-Renyi random graphs. We took $\mathcal{G}_0$ to be an Erdős-Renyi graph on $n = 500$ nodes, with $p = \log(n)/n$. We set the infection parameters to be $k = c = 50$. The alternative graph $\mathcal{G}_1$ was also an Erdős-Renyi random graph, with edges drawn independently with probability p. However, the existence of an edge (u, v) in $\mathcal{G}_0$ and $\mathcal{G}_1$ was correlated, so γp was the probability of (u, v) existing in both graphs, where γ in $[0, 1]$ was taken in increments

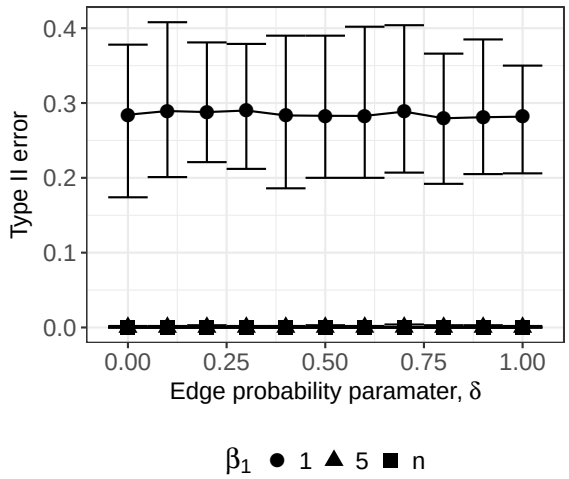

Figure 9.5 The Type II error simulating the distribution of $\mathcal{I}$ on an Erdős-Renyi $\mathcal{G}_0$ with parameters $\beta_0 = n = 500$ against a two-block stochastic block model $\mathcal{G}_1$. Here, $a = 2\delta p$ and $b = 2p - a$ are the within- and between-block edge probabilities. Additionally, we set $k = c = 50$. The marker indicates the average value across 100 graph realizations, and the bars show the range of Type II errors. Perhaps surprisingly, this graph is very similar to Figure 9.4, both in terms of the average Type II error and range of values.

of 0.1. Note that $\gamma = p$ corresponds to independent draws. To estimate thresholds and error probabilities for each graph, we sampled infection processes 1000 times for each set of parameters. Finally, we drew 100 pairs of graphs $(\mathcal{G}_0, \mathcal{G}_1)$.

Due to correlations in edge appearance probabilities in $\mathcal{G}_0$ and $\mathcal{G}_1$, our theory for permutation testing no longer applies. Intuitively, we expect that as γ increases, the test loses its Type I error control: For larger values of γ, the edges of $\mathcal{G}_0$ and $\mathcal{G}_1$ are more positively correlated, which increases the probability that the vertices of an edge (u, v) in $\mathcal{E}_1$ are infected under the null distribution. Thus, the permutation test sets the threshold for the test using W_1 to be too low, causing a large Type I error. Indeed, this behavior is seen in Figure 9.6, where we see that the nominal Type I error level $\alpha = 0.05$ is far from achieved for every realization with $\gamma \geq 0.2$.

The behavior of the Type II error is more nuanced in this case. In particular, for high values of γ, we would still expect the Type II error to be roughly on the same order as in the case of the stochastic block model. However, if we were to set the threshold using the actual null distribution, meaning the correlated random graph $\mathcal{G}_0$, perhaps the threshold would be sufficiently elevated by the correlated edges, which would in turn increase the Type II error. We do indeed see these behaviors in Figures 9.7 and 9.8. In the first figure, the Type II error does not depend on γ, since $\mathcal{G}_1$ is always an Erdős-Renyi random graph with parameter p. In the second figure, we see that the Type II error increases drastically when considering a threshold set by simulating on $\mathcal{G}_0$. However, note that for large values of β_1, the graphs can still be distinguished even with moderately large correlation. This provides hope that in correlated cases, infection patterns that depend very strongly on the graph topology could

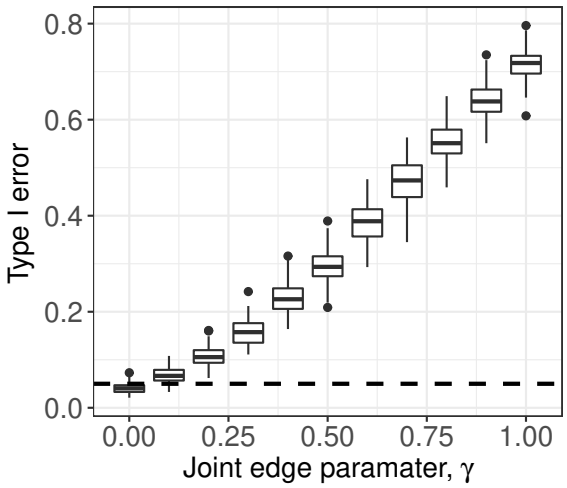

Figure 9.6 The Type I error of the permutation test applied to correlated Erdős-Renyi random graphs on $n = 500$ vertices. The marginal probability of an edge (u, v) is $p = \log(n)/n$ in both the null and alternative, and the probability of the edge in both graphs is γp. As γ increases, the edge correlation increases, causing the Type I error to increase. Here, $\beta = 1$ and $k = c = 50$.

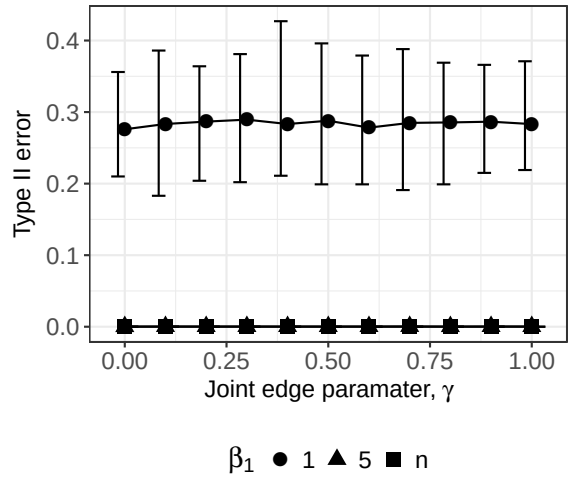

Figure 9.7 The Type II error of the permutation test applied to correlated Erdős-Renyi random graphs on $n = 500$ vertices. The marginal probability of an edge (u, v) is $p = \log(n)/n$ in both the null and alternative, and the probability of the edge in both graphs is γp. The marker indicates the average Type II error, and the bars indicate the range. Perhaps surprisingly, Figure 9.7 resembles Figures 9.4 and 9.5, despite the difference in graph structures.

still yield correct inference for the true graph.

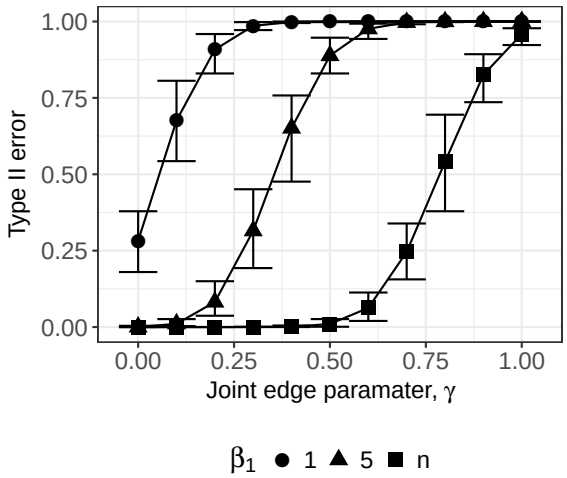

Figure 9.8 The Type II error between a pair of correlated Erdős-Renyi random graphs $\mathcal{G}_0$ and $\mathcal{G}_1$ when $\beta_0 = n = 500$ via null distribution simulation. The marginal probability of an edge (u, v) in either graph is $p = \log(n)/n$, and the probability of the edge in both graphs is γp. Additionally, we set $k = c = 50$. The markers denote the average Type II errors, and the bars denote the range. Note that this figure is very different from Figure 9.7. In particular, for higher values of γ, the Type II error is much worse, since the graphs become similar. However, note that if β_1 is sufficiently high, the graphs can still be distinguished even for moderately large γ.

Bibliography

R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Dynamics and Control. Oxford University Press, 1992.

D. Angluin, J. Aspnes, and L. Reyzin. Network construction with subgraph connectivity constraints. *Journal of Combinatorial Optimization*, 29(2):418–432, 2015.

D. Banerjee. Contiguity and non-reconstruction results for planted partition models: the dense case. *Electronic Journal of Probability*, 23, 2018.

D. Banerjee and Z. Ma. Optimal hypothesis testing for stochastic block models with growing degrees. *arXiv preprint arXiv:1705.05305*, 2017.

C. Borgs, M. Brautbar, J. Chayes, S. Khanna, and B. Lucier. The power of local information in social networks. *Proceedings of the 8th International Workshop on Internet and Network Economics*, pages 406–419, 2012.

S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: a Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

M. Brautbar and M. Kearns. Local algorithms for finding interesting individuals in large networks. *Innovations in Theoretical Computer Science*, 2010.

S. Bubeck, J. Ding, R. Eldan, and M. Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016.

S. Bubeck, L. Devroye, and G. Lugosi. Finding Adam in random growing trees. *Random Structures & Algorithms*, 50(2):158–172, 2017.

W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1029–1038. ACM, 2010.

N. A. Christakis and J. H. Fowler. The spread of obesity in a large social network over 32 years. *The New England Journal of Medicine*, 357(4):370–379, 2007.

R. Cohen, S. Havlin, and D. Ben-Avraham. Efficient immunization strategies for computer networks and populations. *Physical Review Letters*, 91(24):247901, 2003.

T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.

P. T. Darga, M. H. Liffiton, K. A. Sakallah, and I. L. Markov. Exploiting structure in symmetry detection for cnf. In *Proceedings of the 41st Annual Design Automation Conference*, pages 530–534, New York, NY, USA, 2004. ACM. URL http://doi.acm.org/10.1145/996566.996712.

P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 57–66. ACM, 2001.

G. Dudas et al. Virus genomes reveal the factors that spread and sustained the West African Ebola epidemic. *Nature*, 544:309–315, 2017.

D. S. Dummit and R. M. Foote. *Abstract Algebra*. Wiley Hoboken, third edition, 2004.

R. A. Fisher. *The Design of Experiments*. Oliver and Boyd, 1935.

M. Garey and D. Johnson. *Computers and Intractability: A Guide to the Theory of NP-completeness*. Books in mathematical series. W. H. Freeman, 1979. ISBN 9780716710448. URL https://books.google.com/books?id=fjxGAQAAIAAJ.

E. Gilbert and H. Pollak. Steiner minimal trees. *SIAM Journal on Applied Mathematics*, 16 (1):1–29, 1968.

C. Godsil and G. F. Royle. *Algebraic Graph Theory*, volume 207. Springer Science & Business Media, 2013.

M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.

M. Gomez-Rodriguez, L. Song, H. Daneshmand, and B. Schölkopf. Estimating diffusion networks: Recovery conditions, sample complexity and soft-thresholding algorithm. *Journal of Machine Learning Research*, 17(90):1–29, 2016. URL http://jmlr.org/papers/v17/14-430.html.

P. Good. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer Series in Statistics. Springer New York, 2013.

F. Hadlock. Finding a maximum cut of a planar graph in polynomial time. *SIAM Journal on Computing*, 4(3):221–225, 1975.

W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

Y. Huang, M. V. Janardhanan, and L. Reyzin. Network construction with ordered constraints. *arXiv preprint arXiv:1702.07292*, 2017.

M. O. Jackson. *Social and Economic Networks*. Princeton University Press, Princeton, NJ, USA, 2008.

T. Junttila and P. Kaski. Engineering an efficient canonical labeling tool for large and sparse graphs. In *2007 Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments (ALENEX)*, pages 135–149. SIAM, 2007.

R. M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pages 85–103. Springer, 1972.

D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 137–146, New York, NY, USA, 2003. ACM. ISBN 1-58113-737-0.

H. Kesten. On the speed of convergence in first-passage percolation. *The Annals of Applied Probability*, pages 296–338, 1993.

E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer Science & Business Media, 2006.

J. Leskovec and A. Krevl. SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data, 2014.

A. Lubiw. Some NP-complete problems similar to graph isomorphism. *SIAM Journal on Computing*, 10(1):11–21, 1981.

E. M. Luks. Isomorphism of graphs of bounded valence can be tested in polynomial time. *Journal of Computer and System Sciences*, 25(1):42–65, 1982.

B. D. McKay. Computing automorphisms and canonical labellings of graphs. In *Combinatorial Mathematics*, pages 223–232. Springer, 1978.

K. Mehlhorn. A faster approximation algorithm for the steiner problem in graphs. *Information Processing Letters*, 27(3):125–128, 1988.

C. Milling, C. Caramanis, S. Mannor, and S. Shakkottai. Distinguishing infections on different graph topologies. *IEEE Transactions on Information Theory*, 61(6):3100–3120, 2015.

M. Morris. Epidemiology and social networks: modeling structured diffusion. *Sociological Methods and Research*, 22(1):99–126, 1993.

S. A. Myers, C. Zhu, and J. Leskovec. Information diffusion and external influence in networks. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 33–41. ACM, 2012.

P. Netrapalli and S. Sanghavi. Learning the graph of epidemic cascades. In *ACM SIGMETRICS Performance Evaluation Review*, volume 40, pages 211–222. ACM, 2012.

M. E. J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1), 2002.

R. Pastor-Satorras and A. Vespignani. Immunization of complex networks. *Physical Review E*, 65(3):036104, 2002.

J. Potterat, L. Phillips-Plummer, S. Muth, R. Rothenberg, D. Woodhouse, T. Maldonado-Long, H. Zimmerman, and J. Muth. Risk network structure in the early epidemic phase of hiv transmission in colorado springs. *Sexually Transmitted Infections*, 78(suppl 1): i159–i163, 2002.

D. Shah and T. Zaman. Rumors in a network: who's the culprit? *IEEE Transactions on Information Theory*, 2011.

A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009.

Y. Wu. Lecture notes. 2017.