Research Article

Mayleen Cortez-Rodriguez*, Matthew Eichhorn, and Christina Lee Yu

Exploiting neighborhood interference with low-order interactions under unit randomized design

https://doi.org/10.1515/jci-2022-0051
received August 10, 2022; accepted June 22, 2023

Abstract: Network interference, where the outcome of an individual is affected by the treatment assignment of those in their social network, is pervasive in real-world settings. However, it poses a challenge to estimating causal effects. We consider the task of estimating the total treatment effect (TTE), or the difference between the average outcomes of the population when everyone is treated versus when no one is, under network interference. Under a Bernoulli randomized design, we provide an unbiased estimator for the TTE when network interference effects are constrained to low-order interactions among neighbors of an individual. We make no assumptions on the graph other than bounded degree, allowing for well-connected networks that may not be easily clustered. We derive a bound on the variance of our estimator and show in simulated experiments that it performs well compared with standard estimators for the TTE. We also derive a minimax lower bound on the mean squared error of our estimator, which suggests that the difficulty of estimation can be characterized by the degree of interactions in the potential outcomes model. We also prove that our estimator is asymptotically normal under boundedness conditions on the network degree and potential outcomes model. Central to our contribution is a new framework for balancing model flexibility and statistical complexity as captured by this *low-order interactions* structure.

Keywords: causal inference, total treatment effect, neighborhood interference, design of experiments

MSC 2020: 62K99, 91D30, 60F05

1 Introduction

Accurately estimating causal effects is relevant in numerous applications, from pharmaceutical companies researching the efficacy of a new medication, to policy makers understanding the impact of social welfare programs, to social media companies evaluating the impact of different recommendation algorithms on user engagement across their platforms. Often, the entity interested in understanding a causal effect will design an experiment where they randomly assign subsets of the population to treatment (e.g., new medication) and to control (e.g., a placebo) and draw conclusions based on the observed outcomes of the participants (e.g., health outcomes). Other times, the entity may need to determine causal effects from observational data accrued in a previous study where they did not have full control over the treatment assignment mechanism.

*** Corresponding author: Mayleen Cortez-Rodriguez,** Center for Applied Mathematics, Cornell University, Ithaca NY, 14850, USA, e-mail: mec383@cornell.edu
Matthew Eichhorn: Center for Applied Mathematics, Cornell University, Ithaca NY, 14850, USA, e-mail: meichhorn@cornell.edu
Christina Lee Yu: Department of Operations Research and Information Engineering, Cornell University, Ithaca NY, 14850, USA, e-mail: cleeyu@cornell.edu

Our work focuses on estimating the *total treatment effect* (TTE), or the difference between the average outcomes of the population when everyone is treated versus when no one is treated, given data collected from a *randomized experiment*. This estimand is sometimes referred to as the global average treatment effect, as in ref. [1]. The TTE is a quantity of interest in scenarios where the decision maker must choose between adopting the proposed treatment or sticking with the *status quo*. For example, a social media company may develop a new recommendation algorithm for suggesting content to their users, and they want to decide whether to roll out this new algorithm across their platform. As another example, suppose that a pharmaceutical company develops a vaccine for some infectious disease. Then, public health experts and officials must decide whether the vaccine is safe and efficacious enough to warrant its recommendation to the general population. As the side effects of the new treatment are unknown, the goal is to determine the efficacy of the treatment relative to the status quo baseline by running a budgeted randomized trial, where the number of treated individuals in the trial is limited for safety reasons.

The techniques and guarantees for estimating causal effects in classical causal inference heavily rely upon the stable unit treatment value assumption (SUTVA), which posits that the outcome of each individual is independent of the treatment assignment of all other individuals [2]. Unfortunately, SUTVA is violated in all the aforementioned applications because people influence and are impacted by their peers. In the presence of this *network interference*, an individual's outcome is affected by the treatment assignment of others in their social network, and SUTVA no longer holds. Distinguishing between the direct effect of treatment on an individual and the network effect of others' treatment on the individual can be challenging. This has resulted in a growing literature on causal inference in the presence of network (interference) effects, sometimes referred to as spillover or peer influence effects. In this work, we consider the task of estimating the TTE from unit randomized trials under *neighborhood interference*, when an individual is affected by the treatment of its direct neighbors but is otherwise unaffected by the treatment of the individuals outside their neighborhood. Furthermore, we focus on *unit* randomized designs (RDs), wherein individuals are independently assigned to either treatment or control in a randomized experiment. This is in contrast to cluster RDs, which have been proposed as an approach to address network interference for randomized experimental design but which may not be feasible in practice due to an incompatibility with existing experimental platforms.

Estimating a causal effect from randomized experiments involves two decisions. First, we must decide what kind of experiment to run, i.e., how we choose the individuals to participate in the study, and how we determine which individuals are assigned to receive the new treatment. In this article, we focus on randomization inference, in which we assume the entire population participates in the study, and the randomness arises from the assignment of treatment or control. Second, after the experiment is conducted, we must decide how to analyze the data and construct an estimate from the measured observations. The literature addressing network interference can largely be categorized as either constructing clever RDs to exploit clustering structure in the network, or designing new estimators that exploit structure in the potential outcomes model. Without making assumptions on either the network or the potential outcomes model, it is impossible to estimate the TTE under arbitrary interference [3–5].

Table 1 summarizes how different assumptions on the potential outcomes model or network structure lead to different types of solutions, with a focus on unbiased estimators and neighborhood interference models where the network is known. The columns correspond to assumptions on the network structure (in order of increasing generality from left to right), while the rows correspond to assumptions on the structure in the potential outcomes model (in order of increasing generality from top to bottom). The body of the graph lists proposed solutions for estimating the TTE under the corresponding assumptions on the network/model structure. For example, under a fully general network structure and a linear potential outcomes model, refs [6–11] propose using an ordinary least squares (OLS) estimator with a Bernoulli RD. Since we focus on solutions proposing *unbiased* estimators, we also list bounds on the variance of the estimators in the table, as a point of comparison. As Table 1 indicates, the literature has either focused on analyzing the Horvitz–Thompson estimator under new RDs that exploit network structure by increasing the correlation between neighbors' treatment assignments or alternatively using regression-style estimators with Bernoulli RD, exploiting strong functional assumptions on the potential outcomes model. Without assuming any structure

Table 1: Literature Summary. Each row corresponds to an assumption on the structure of the potential outcomes model, and each column corresponds to an assumption on the structure of network. We list a proposed solution under the corresponding model/network assumptions in the following order: an unbiased estimator for the TTE, the RD, a bound on the variance (if available), and citations to related work. In the variance bounds, $Y_{\max}$ is a bound on the effects on any individual, d is the maximum neighborhood size, C is the number of subcommunities or clusters, p is the treatment probability that is assumed to be small, κ is the restricted growth parameter, and β is the polynomial degree of the potential outcomes model. Our result proposes an estimator for the TTE under a β-order interactions (equivalently, β-degree polynomial) structure, a fully general graph, and Bernoulli design. All the solutions in the table rely on full knowledge of the network under neighborhood interference, and we focus on unbiased estimators.

Model structure	**Network structure**		
	C disconnected subcommunities	**κ-restricted growth**	**Fully general**
Linear **Generalized linear**	Directions for future work		OLS, *Bernoulli RD*; [6–11] Regression/machine learning, *Bernoulli RD*; [11]
β-order interactions			SNIPE, *Bernoulli RD*; $O\left(\frac{Y_{\max}^2 d^{2\beta+2}}{np^\beta}\right)$
Arbitrary neighborhood interference	*Horvitz–Thompson*, Cluster RD; $O\left(\frac{Y_{\max}^2}{Cp}\right)$; [12–15]	*Horvitz–Thompson*, Randomized Cluster RD; $O\left(\frac{Y_{\max}^2 \kappa^4 d^2}{np}\right)$; [1,7,16,17]	*Horvitz–Thompson*, *Bernoulli RD*; $O\left(\frac{Y_{\max}^2 d^2}{np^d}\right)$; [3]

beyond neighborhood interference, the baseline solution of using the Horvitz–Thompson estimator under the Bernoulli RD is an unbiased estimator whose variance scales exponentially in the degree of the network.

In our work, we propose a hierarchy of model classes that extrapolates between simple linear models and complex general models, such that a practitioner can choose the strength of the model assumptions they are willing to impose. Naturally, assuming a more limited model simplifies the causal inference task. We characterize the complexity of a potential outcomes model with the order of interactions β, which also corresponds to the polynomial degree of the potential outcomes function when viewed as a polynomial of the treatment vector. A β-order interactions model is one in which each neighborhood set of size at most β can have a unique additive network effect on the potential outcome. Our model allows for heterogeneity in the influence of different sets of treated neighbors, strictly generalizing beyond the typical parametric model classes used in the literature, which oftentimes assumes anonymous interference. We make no assumptions on the graph beyond bounded degree, so the graph may be well connected and not easily clustered. We summarize our contributions and results below:

(1) Assuming a β-order interactions model, under a nonuniform Bernoulli RD with treatment probabilities $\{p_i\}_{i=1}^n$ for $p_i \in [p, 1-p]$, we present the structured neighborhood interference polynomial estimator (SNIPE), a simple unbiased estimator for the TTE.
(2) We derive a bound on the variance of our estimator, which scales polynomially in the degree d of the network and exponentially in the degree β of the potential outcomes model. We also show that our estimator is asymptotically normal.
(3) For a d-regular graph and uniform treatment probabilities $p_i = p$ with $p^\beta < 0.16$ and $\beta \ll d$, we prove that the minimax optimal mean squared error for estimating the TTE is lower bounded by $\Omega\left(\frac{1}{np^\beta}\right)$, implying that the exponential dependence on β is necessary.
(4) We provide experimental results to illustrate that using regression models that do not allow for heterogeneity among the network effects can lead to considerable bias when the anonymous interference assumption is violated. The experiments validate that our estimator is unbiased for β-order interaction models, and obtains a lower mean squared error than existing alternatives.

To interpret the upper bound on the variance of our proposed estimator for the TTE, we note that our variance scales as $O(\text{poly}(d)/np^\beta)$, compared to the variance of the Horvitz–Thompson estimator under

Bernoulli RD which scales as $O(1/np^d)$, where β is always bounded above by d. For smaller values of β, the β-order interactions model imposes stronger structural assumptions on the potential outcomes model than required by the Horvitz–Thompson estimator. In turn, our estimator has significantly lower variance, scaling only polynomially in the network degree d yet exponential in β, as opposed to exponential in d. In addition, the minimax lower bound shows that the exponential dependence on β is also minimax optimal amongst β-order interaction models, implying that the order of interactions, or polynomial degree, of the potential outcomes model is a meaningful property that expresses the complexity of estimating the TTE.

2 Related literature

While there has been a flurry of recent activity in addressing the challenges that arise from network interference, every proposed solution concept fundamentally hinges upon making key assumptions on the form of network interference. Without any assumptions, the vector of observed outcomes under a particular treatment vector $\mathbf{z} \in \{0,1\}^n$ may have no relationship to the potential outcomes under any other treatment vector. We should naturally expect that if we are willing to assume stronger assumptions, then we may be able to obtain stronger results conditioned on those assumptions being satisfied. As such, the literature ranges between results that impose fewer assumptions on the model and graph, resulting in unbiased estimators with high variance, or results that impose strong assumptions on the model and graph, resulting in simple, unbiased estimators with low variance.

Early works propose framing assumptions via exposure functions, or constant treatment response [3,4]. This assumes that there is some known exposure mapping, $f(\mathbf{z}, \theta_i) \in \Delta$, which maps the treatment vector $\mathbf{z}$, along with unit-specific traits θ_i, to a discrete space Δ of *exposures*, or effective treatments. The potential outcomes function for unit i is then assumed to only be a function of its *exposure*, or effective treatment, such that if $f(\mathbf{z}, \theta_i) = f(\mathbf{z}', \theta_i)$, then $Y_i(\mathbf{z}) = Y_i(\mathbf{z}')$. If $|\Delta|$ is as large as 2^n, then this assumption places no effective restriction on the potential outcomes function; thus, this assumption is only meaningful when $|\Delta|$ is relatively small. One commonly used exposure mapping expresses the neighborhood interference assumption [17–20], in which each unit i is associated to a neighborhood $\mathcal{N}_i \subseteq [n]$, and unit i's potential outcome is only a function of the treatments of units within $\mathcal{N}_i$. We could use an exposure mapping to formulate this assumption, where $|\Delta| = 2^d$ for d denoting the maximum neighborhood size, and $f(\mathbf{z}, \mathcal{N}_i) = f(\mathbf{z}', \mathcal{N}_i)$ if the treatments assigned to individuals in $\mathcal{N}_i$ are the same between $\mathbf{z}$ and $\mathbf{z}'$. Another commonly used exposure mapping expresses the anonymous interference assumption [14,21–23], in which the potential outcomes are only a function of the treatments through the number of treated neighbors, i.e., $f(\mathbf{z}, \mathcal{N}_i) = f(\mathbf{z}', \mathcal{N}_i)$, if the number of treated individuals in $\mathcal{N}_i$ via $\mathbf{z}$ and $\mathbf{z}'$ is the same. While the exposure mapping framework provides a highly generalizable and abstract framework for assumptions, it is fundamentally discrete in nature and the complexity of estimation is characterized by the number of possible exposures $|\Delta|$, which could still be large. As a result, ref. [4] suggests a collection of additional assumptions that can be imposed on top of anonymous neighborhood interference, including distributional or functional form assumptions, or additivity assumptions as suggested in ref. [18].

The majority of works in the literature (along with our work) assume neighborhood interference with respect to a known graph. Notable exceptions include ref. [24], which considers a highly structured *spatial interference* setting with network effects decaying with distance, and [25], which provides methods for testing hypotheses about interference effects including higher order spillovers. Without imposing any additional assumptions on the potential outcomes besides neighborhood interference, a natural solution is to use the Horvitz–Thompson estimator to estimate the average outcomes under full neighborhood treatment and full neighborhood control [3]. While the estimator is unbiased, the variance of the estimator scales inversely with the probability that a unit's full neighborhood is in either treatment or control. Under a Bernoulli(p) RD, where each individual is treated independently with probability p, the variance scales as $O(1/np^d)$, as indicated in the bottom right cell of Table 1. The exponential dependence on d renders the estimator impractical for realistic networks.

One approach in response to the aforementioned challenge is to consider cleverly constructing RDs that increase overlap, i.e., the probability that a unit's entire neighborhood is assigned to treatment or control. The earliest literature in this line of work additionally assumes *partial interference*, also referred to as the multiple networks setting, in which the population can be partitioned into disjoint groups, and network interference only occurs within groups and not across groups [12,14,15,20,21,26,27]. This assumption makes sense in contexts where interference is only expected between naturally clustered groups of individuals, such as households, cities, or countries. Given knowledge of the groups, we can then randomly assign groups to different treatment saturation levels, e.g., jointly assigning groups to either treatment or control, increasing the correlation of treatments within neighborhoods. Then, a difference in means estimator or a Horvitz–Thompson estimator can be used to estimate the TTE. The asymptotic consistency of these estimators relies on the number of groups going to infinity, with a variance scaling inversely with the fraction of treated clusters, i.e., $O(1/Cp)$ as indicated in the bottom left cell of Table 1. In practice, even networks that are clearly clustered into separate groups may not have a sufficiently large number of groups to result in accurate estimates.

For general connected graphs, one can still implement a cluster-based RD on constructed clusters, where the clusters are constructed to minimize the number of edges between clusters [1,7,16,17]. References [1,17] provide guarantees for graphs exhibiting a restricted growth property, which limits the rate at which local neighborhoods of the graph can grow in size, and ref. [1] proves that using randomized cluster RD along with the Horvitz–Thompson estimator achieves a variance of $O(1/np)$ for restricted growth networks, which is a significant gain from the exponential dependence on d under Bernoulli design. A limitation of these solutions is that the cluster RD can be difficult to implement due to incompatibility with existing experimentation platforms or to ethical concerns. When the existing experimentation platform is already set up for a unit RD, the experimenter may have the desire to avoid overhauling the platform to work with cluster RD due to time or resource constraints [11]. In addition, refs [28–31] detail some ethical issues that arise in cluster RDs including, but not limited to, problems with informed consent (e.g., it may be infeasible to gain informed consent from every unit in a cluster) and concerns about distributional justice (e.g., with regards to how clusters are selected and assigned to treatment). Furthermore, both refs [28,30] comment that many of the existing, formal research ethics guidelines were designed with unit RD in mind and thus, offer little guidance for considerations that arise in cluster RDs. The work we present here focuses on scenarios with unit RDs, when cluster RDs may not be feasible or may be undesirable due to any of the aforementioned concerns.

The alternate approach that has gained traction empirically is to impose additional functional assumptions on the potential outcomes in addition to anonymous neighborhood interference. The most common assumption is that the potential outcomes are linear with respect to a particular statistic of the treatment vector, where the linear function is shared across units [6–11]. Under this assumption, estimating the entire potential outcomes function reduces to linear regression, which one could solve using OLS, as indicated in the upper right most cell of Table 1. After recovering the linear model, one could estimate any desired causal estimand. The results rely on correctly choosing the statistic that governs the linearity, or more generally reduces the task to heuristic *feature engineering* for generalized linear models [11]. One could plausibly extend the function class beyond linear and instead apply machine learning techniques to estimate the function that generalizes beyond linear regression. While we do not state a variance bound in Table 1, one would expect that when p is sufficiently large, the variance would scale with $O(1/n)$, with some dependence on the complexity of the model class; when p is very small, the variance may scale with $O(1/pn)$, as the regression still requires sufficient variance of covariates represented in the data, i.e., a sufficient spread of number of neighbors treated. Reference [22] considers nonparametric models yet focuses on estimating a locally linearized estimand. A drawback of these approaches is that they assume *anonymous interference*, which imposes a symmetry in the potential outcomes such that the causal network effect of treating any of an individual's neighbors is equal regardless of which neighbor is chosen. In addition, they assume that the function that we are learning is shared across units, or at least units of the same known covariate type, which can be limiting.

While we have primarily focused on summarizing the literature for unbiased estimators, there has also been some work considering how structure in the potential outcomes model can be exploited to reduce bias of standard estimators. References [32–34] use application domain-specific structure in two-sided marketplaces

and network routing to compare the bias of the difference in means estimators under different experimental designs. Reference [35] uses structure in ridesharing platforms to propose a new estimator with reduced bias. In addition, there has been some limited empirical work studying bias under model misspecification [16].

Complementary to the literature on randomized experiments, there has been a growing literature considering causal inference over observational studies in the presence of network interference. However, the limitations similarly mirror the aforementioned concerns. A majority of the literature assumes partial interference [15,36–39], i.e., the multiple networks setting, which then enables causal inference of a variety of different estimands. In particular, it is commonly assumed that the different groups in the network are sampled from some underlying distribution, and the statistical guarantees are given with respect to the number of groups going to infinity. Alternately, other works assume that the potential outcomes only depend on a simple and known statistic of the neighborhood treatment, most commonly the number or fraction of treated [11,40,41]. Either the neighborhood statistic only takes finite values, or assumptions are imposed on the functional form of the potential outcomes, which imply anonymous interference and reduce inference to a regression task or maximum likelihood calculation. Reference [42] considers a general exposure mapping model alongside an inverse propensity weighted estimator, but the estimator has high variance when the exposure mapping is complex.

In contrast to the majority of the mentioned literature, we neither rely on cluster RDs nor anonymous interference assumptions. We instead propose a potential outcomes model with *low-order interactions* structure, where the degree of interactions β characterizes the difficulty of inference, also studied in ref. [43]. For $\beta = 1$, this model is equivalent to heterogeneous additive network effects in ref. [44], which can be derived from the joint assumptions of additivity of main effects and interference in ref. [18]. When β is larger than the network degree, then this assumption is equivalent to an arbitrary neighborhood interference, providing a nested hierarchy of models that encompass the simple linear model class, the fully general model class, as well as model classes of varying complexity in between. References [43,44], which consider similar models as we do, focus on the setting when the underlying network is fully unknown, and yet there is richer available information either in the form of historical data [44] or multiple measurements over the course of a multi-stage experiment [43], neither of which we assume in this work. Our estimator also has close connections to the pseudoinverse estimator in ref. [45], the Riesz estimator in ref. [46], and the estimator in ref. [22], which is discussed in Section 4.

3 Model

3.1 Causal network

Let $[n] \coloneqq \{1, \ldots, n\}$ denote the underlying population of n individuals. We model the network effects in the population as a directed graph over the individuals with edge set $E \subseteq [n] \times [n]$. An edge $(j, i) \in E$ signifies that the treatment assignment of individual j affects the outcome of individual i. As an individual's own treatment is likely to affect their outcome, we expect self-loops in this graph. In much of the article, we are concerned with neighborhood interference effects, and we use $\mathcal{N}_i \coloneqq \{j \in [n] : (j, i) \in E\}$ to denote the in-neighborhood of an individual i. Note that this definition allows $i \in \mathcal{N}_i$. Many of our variance bounds reference the network degree. We let d_{in} denote the maximum in-degree of any individual, d_{out} denote the maximum out-degree, and $d \coloneqq \max\{d_{\text{in}}, d_{\text{out}}\}$.

3.2 Potential outcomes model

To each individual i, we associate a treatment assignment $z_i \in \{0, 1\}$, where we interpret $z_i = 1$ as an assignment to the treatment group and $z_i = 0$ as an assignment to the control group. We collect all treatment

assignments into the vector $\mathbf{z}$. We use Y_i to denote the outcome of individual i. As our setting assumes network interference, the classical SUTVA assumption is violated. That is, Y_i is not a function only of z_i. Rather, $Y_i : \{0,1\}^n \to \mathbb{R}$ may be a function of $\mathbf{z}$, the treatment assignments of the entire population. Since each treatment variable z_i is binary, we can indicate an exact treatment assignment as a product of z_i (for treated individuals) and $(1 - z_i)$ (for untreated individuals) factors. As such, we can express a general potential outcome function Y_i as a polynomial in $\mathbf{z}$,

$$Y_i(\mathbf{z}) = \sum_{\mathcal{T} \subseteq [n]} a_{i,\mathcal{T}} \prod_{j \in \mathcal{T}} z_j \prod_{k \in [n] \setminus \mathcal{T}} (1 - z_k),$$

where $a_{i,\mathcal{T}}$ is individual i's outcome when their set of treated neighbors is exactly $\mathcal{T}$. Via a change of basis, we can equivalently express $Y_i(\mathbf{z})$ as a polynomial in the "treated subsets":

$$Y_i(\mathbf{z}) = \sum_{\mathcal{S}' \subseteq [n]} c_{i,\mathcal{S}'} \prod_{j' \in \mathcal{S}'} z_{j'}, \tag{3.1}$$

where $c_{i,\mathcal{S}'}$ represents the additive effect on individual i's outcome that they receive when the entirety of subset $\mathcal{S}'$ (perhaps among other individuals) is treated. Note that $c_{i,\varnothing}$ represents the *baseline effect*, the component of i's outcome that is independent of the treatment assignments.

So far, the potential outcomes model described in (3.1) is completely general. However, it is parameterized by 2^n coefficients $\{c_{i,\mathcal{S}'}\}$, which makes it untenable in most settings. To combat this, we impose some structural assumptions on these coefficients. First, we observe that the populations of interest can be quite large (e.g., the population of an entire country), and their influence networks may have high diameter. Throughout most of the article, we assume that individuals' outcomes are influenced only by their immediate in-neighborhood.

Assumption 1. (Neighborhood interference) $Y_i(\mathbf{z})$ only depends on the treatment of individuals in $\mathcal{N}_i$. Equivalently, $Y_i(\mathbf{z}) = Y_i(\mathbf{z}')$ for any $\mathbf{z}$ and $\mathbf{z}'$ such that $z_j = z'$ for all $j \in \mathcal{N}_i$. In our notation, $c_{i,\mathcal{S}'} = 0$ for any $\mathcal{S}' \not\subseteq \mathcal{N}_i$.

Next, we note that the degree of each $Y_i(\mathbf{z})$ can (under the neighborhood interference assumption) be as large as d_{in}. In such a model, one's outcome may be differently influenced by a treated coalition of any size in their neighborhood. Contrast this with a simpler linear potential outcomes model, wherein an individual's outcome receives only an independent additive effect from each of their treated neighbors. This illustrates that the degree of the polynomial may serve as a proxy for its complexity. In this work, we consider the scenario where the polynomial degree may be significantly smaller than d_{in}.

Assumption 2. (Low polynomial degree) Each potential outcome function $Y_i(\mathbf{z})$ has degree at most β. In our notation, $c_{i,\mathcal{S}'} = 0$ whenever $|\mathcal{S}'| > \beta$. Along with Assumption 1, it follows that the potential outcomes function $Y_i(\mathbf{z})$ from (3.1) can be expressed in the following form,

$$Y_i(\mathbf{z}) = \sum_{\mathcal{S}' \in \mathcal{S}_i^\beta} c_{i,\mathcal{S}'} \prod_{j \in \mathcal{S}'} z_j, \tag{3.2}$$

where we define $\mathcal{S}_i^\beta \coloneqq \{\mathcal{S} \subseteq \mathcal{N}_i : |\mathcal{S}| \leq \beta\}$.

We remark that while we use the formal mathematical term of "low polynomial degree," since this describes a function over a vector of binary variables, a low polynomial degree constraint is equivalent to a constraint on the *order of interactions* amongst the treatments of neighbors. In the simplest setting when $\beta = 1$, this is equivalent to a model in which the networks effects are additive across treated neighbors, strictly generalizing beyond all linear models that have been widely used in applied settings.

We use the notation in (3.2) to express the potential outcomes model for the remainder of the article. If $\beta \geq d_{\text{in}}$, note that β could be completely removed from the definition of Y_i in equation (3.2), reducing to the arbitrary neighborhood interference setting. However, we turn our attention to settings where β might be much smaller than the degree of the graph ($\beta \ll d_{\text{in}}$), and we can assume that only low-order interactions within neighborhoods have an effect on an individual. As noted earlier, taking $\beta = 1$ corresponds to the

heterogeneous linear outcomes model in ref. [44]. We include further examples to help in understanding this low polynomial degree assumption in Section 3.2.1.

Many of our variance bounds utilize an upper bound on the treatment effects for each individual. We define $Y_{\max}$ such that

$$Y_{\max} \coloneqq \max_{i\in[n]} \sum_{\mathcal{S}'\subseteq\mathcal{S}_i^{\beta}} |c_{i,\mathcal{S}'}|. \tag{3.3}$$

It follows that $|Y_i(\mathbf{z})| \le Y_{\max}$ for any treatment vector $\mathbf{z}$.

Remark 1. We emphasize that the model in (3.2) captures fully general neighborhood interference when $\beta = d$. Even when $\beta < d$, the number of parameters in the model grows with n. This results in the total number of observations always being smaller than the number of parameters in the model, so using simple regression-style estimators to identify the model is impossible.

Remark 2. We can consider the order of interactions or polynomial degree β as a way to measure the complexity of the model class. Our subsequent results suggest that β is meaningful as it also captures a notion of statistical complexity or difficulty of estimation. This is evidenced by the variance upper bound and minimax lower bound results in Section 5, where we see that the smaller β is relative to the degree of the graph, the smaller the variance incurred and the larger β is with respect to the graph degree, the higher the variance incurred. In some sense, the "lower" the degree of the model, i.e., the more structure imposed, the "easier" it is to estimate the TTE. On the flip side, the "higher" the degree of the model, i.e., the closer it is to being fully general, the "harder" it is to estimate.

Remark 3. Assumption 2 implies that the model exhibits a particular type of sparsity in the coefficients with respect to the monomial basis, in which the coefficients corresponding to sets larger than size β are zero. In our setting when the treatments are budgeted, i.e., p is small, these coefficients precisely correspond to effects that are observable in a Bernoulli RD, i.e., the probability for observing or measuring the coefficient corresponding to set $\mathcal{S}$ is $p^{|\mathcal{S}|}$. As such, there is an direct connection between this hierarchy of models as parameterized by β and the ability to measure and estimate the TTE, which is formalized by our subsequent analysis.

3.2.1 Examples of low-degree interaction models

We provide a few examples to illustrate when the polynomial degree of the potential outcomes model may naturally be smaller than the total neighborhood size (i.e., $\beta < d$).

Example 1. Consider a potential outcomes model that exhibits the joint assumptions of additivity of main effects and interference effects as defined in ref. [18]. This imposes that the potential outcomes satisfy

$$Y_i(\mathbf{z}) = Y_i(\mathbf{0}) + (Y_i(z_i\mathbf{e}_i) - Y_i(\mathbf{0})) + \sum_{k\in[n]} (Y_i(z_k\mathbf{e}_k) - Y_i(\mathbf{0})),$$

where $\mathbf{0} \in \mathbb{R}^n$ is a vector of all zeros and $\mathbf{e}_j \in \{0,1\}^n$ is a standard basis vector. As discussed in ref. [44], this assumption implies a low-degree interaction model with $\beta = 1$, which they refer to as heterogeneous additive network effects.

Example 2. Consider a hypothesized setting where each individual's neighborhood can be divided into smaller *sub-communities*. For example, one's close contacts may include their immediate family, their close friends, their work colleagues, etc. Within each of these sub-communities, there may be nontrivial (higher order) interactions between the treatments of its members. However, it is reasonable to assume that the cumulative effects of each sub-community have an additive effect on the individual's outcome. That is, there are no

nontrivial interactions resulting from the treatment of individuals across different sub-communities. In this case, a natural choice for β is the size of the largest sub-community, which could be significantly smaller than the size of the largest neighborhood.

Example 3. Suppose a social networking platform is testing a new "hangout room" feature where groups of up to five people can interact in a novel environment on the platform. One can posit that a natural choice for β in this setting is 5, as the change in any individual's usage on the platform can be attributed to how they utilize this new feature, which in turn is a function of it being introduced to various subsets of up to five users in that individual's neighborhood.

Example 4. Consider a setting where an individual's outcome is a low-degree polynomial in some auxiliary quantity, which is itself linear in the treatment assignments of their neighborhood. For example, we might have

$$Y_i(\mathbf{z}) = c_{i,\varnothing} + \sum_{j \in \mathcal{N}_i} c_{ij} z_j + \left(\sum_{j \in \mathcal{N}_i} c_{ij} z_j \right)^2 + \cdots + \left(\sum_{j \in \mathcal{N}_i} c_{ij} z_j \right)^\beta.$$

A similar setting is explored in our simulated experiments in Section 6.

Example 5. Consider an example in which network effects only arise from pairwise edge interactions and triangular interactions, i.e., individual i's outcome consists of a sum of its baseline outcome, its own direct effect $c_{i,i}$, pairwise edge effects $c_{i,j}$ for $j \in \mathcal{N}_i$, and triangle effects $c_{i,\{j,j'\}}$ for $j, j' \in \mathcal{N}_i$ such that there is also an edge between j and j', indicating that the three individuals are mutual connections. For such a model, β would be 2 due to the triangular interactions.

Remark 4. We emphasize that any potential outcomes model that takes a binary treatment vector $\mathbf{z}$ as its input can be written as a polynomial in $\mathbf{z}$, taking the form in equation (3.2) for general β. However, the low-degree assumption, i.e., that $\beta \ll d$, will not generally admit high-degree models. For example, both threshold models and saturation models generally require the degree of $Y_i(\mathbf{z})$ to be $|\mathcal{N}_i|$. In a threshold potential outcomes model, an individual experiences network effects once a particular threshold of their neighbors are treated [7]. An example of this type of model is given by

$$Y_i(\mathbf{z}) = \mathbb{I}\left(\sum_{j \in \mathcal{N}_i} z_j \geq \theta \right) \cdot \sum_{\mathcal{S} \subseteq \mathcal{N}_i} c_{i,\mathcal{S}} \prod_{j \in \mathcal{S}} z_j,$$

where $0 \leq \theta \leq |\mathcal{N}_i|$. Saturation models allow for network or peer effects to increase up to a particular saturation level, such as

$$Y_i(\mathbf{z}) = \min\left(\theta, \sum_{\mathcal{S} \subseteq \mathcal{N}_i} c_{i,\mathcal{S}} \prod_{j \in \mathcal{S}} z_j \right),$$

where θ is some maximum saturation threshold. In this model, an individual receives additive effects from each subset of treated neighbors until a certain threshold effect is reached, after which the networks effects have "saturated" and the treatment of additional neighbors contributes no additional effect.

3.3 Causal estimand and RD

Throughout most of the article, we concern ourselves with estimating the TTE. This quantifies the difference between the average of individual's outcomes when the entire population is treated versus the average of individual's outcomes when the entire population is untreated:

$$\text{TTE} \coloneqq \frac{1}{n}\sum_{i=1}^{n}(Y_i(\mathbf{1}) - Y_i(\mathbf{0})), \tag{3.4}$$

where $\mathbf{1}$ represents the all ones vector and $\mathbf{0}$ represents the zero vector. Plugging in our parameterization from equation (3.2), we may re-express the TTE as follows:

$$\text{TTE} = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}'\in\mathcal{S}_i^{\beta}\\ \mathcal{S}'\neq\varnothing}} c_{i,\mathcal{S}'}. \tag{3.5}$$

Since exposing individuals to treatment can have a deleterious and irreversible effect on their outcomes, we wish to estimate the TTE after treating a small random subset of the population. We focus on a *nonuniform* Bernoulli design, wherein each individual i is independently assigned treatment with probability $p_i \in [p, 1-p]$ for $p > 0$. That is, each $z_i \sim \text{Bernoulli}(p_i)$. Such a RD is both straightforward to implement and to understand. Furthermore, many existing experimentation platforms are already designed for Bernoulli randomization, making it easy to collect new data or to re-analyze existing data and adjust for network interference, rather than requiring a complete overhaul of the existing experimentation platform to allow for more complex randomization schemes.

4 Estimator

In this section, we introduce the estimator that will serve as our central focus through the rest of the article: the SNIPE. While we restrict our attention to nonuniform Bernoulli design, the techniques to derive this estimator can be generalized to a wide variety of causal estimands and experimental designs. In fact, our estimator is connected to the Horvitz–Thompson estimator and turns out to be a special case of both the *pseudoinverse estimator* first introduced by Swaminathan et al. [45] and more recently the *Riesz estimator* of Harshaw et al. [46]. We discuss these connections in Section 4.3.

To provide intuition, we first derive the estimator in the $\beta = 1$ case with the linear heterogenous potential outcomes model [44] via a connection to OLS. Then, we show how it can be extended to the more general polynomial setting. Our main result (Section 5) establishes both the unbiasedness of this family of estimators and a bound on its variance under Bernoulli design.

Recalling that $\mathcal{S}_i^{\beta} \coloneqq \{\mathcal{S} \subseteq \mathcal{N}_i : |\mathcal{S}| \leq \beta\}$, the vector $\mathbf{c}_i$ collects the parameters $\{c_{i,\mathcal{S}}\}_{\mathcal{S}\in\mathcal{S}_i^{\beta}}$ in some canonical ordering. We will assume throughout that $\varnothing \in \mathcal{S}_i^{\beta}$ is always first in this ordering and otherwise index the entries of these vectors by their corresponding set $\mathcal{S}$. As an example, when $\beta = 1$ and $\mathcal{N}_i = \{j_1, \ldots, j_{d_i}\}$, we may have $\mathbf{c}_i = [c_{i,\varnothing}\, c_{i,\{j_1\}} \ldots c_{i,\{j_{d_i}\}}]^{\mathsf{T}}$, where d_i is the in-degree of unit i. In a similar manner, the *treated subsets vector* $\tilde{\mathbf{z}}_i$ collects the indicators $\{\prod_{j\in\mathcal{S}} z_j\}_{\mathcal{S}\in\mathcal{S}_i^{\beta}}$ in the same ordering. In our $\beta = 1$ example, $\tilde{\mathbf{z}}_i = [1\, z_{j_1} \ldots z_{j_{d_i}}]^{\mathsf{T}}$. By using this notation, we may express

$$\mathbf{Y}_i(\mathbf{z}) = \langle \mathbf{c}_i, \tilde{\mathbf{z}}_i\rangle,\ \text{TTE} = \frac{1}{n}\sum_{i=1}^{n}\langle(\mathbf{1}_{|S_i|} - \mathbf{e}_1), \mathbf{c}_i\rangle, \tag{4.1}$$

where the inner product argument in the second equation is the $|S_i|$-length vector with first entry 0 and remaining entries 1.

4.1 Building intuition in the linear setting ($\beta = 1$)

To motivate our estimator, we consider the linear heterogeneous potential outcomes model ($\beta = 1$) under nonuniform Bernoulli RD. By using the TTE expression from (4.1), we can recast the problem of estimating the TTE into the problem of estimating the parameter vector $\mathbf{c}_i$: by linearity of expectation, an unbiased estimator of $\mathbf{c}_i$ will give rise to an unbiased estimator of TTE.

As a thought experiment toward estimating $\mathbf{c}_i$, imagine that we can perform M independent replications of our randomized experiment. In each replication $m \in [M]$, we observe the treated subsets vector $(\tilde{\mathbf{z}}_i)^m$, obtained from our Bernoulli RD, and the outcome $(Y_i)^m$. We visualize,

$$\underbrace{\begin{bmatrix}(Y_i)^1\\(Y_i)^2\\\vdots\\(Y_i)^M\end{bmatrix}}_{\mathbf{Y}_i\in\mathbb{R}^M} = \underbrace{\begin{bmatrix}1 & z_{j_1}^1 & \dots & z_{j_{d_i}}^1\\1 & z_{j_1}^2 & \dots & z_{j_{d_i}}^2\\\vdots & \vdots & \ddots & \vdots\\1 & z_{j_1}^M & \dots & z_{j_{d_i}}^M\end{bmatrix}}_{\mathbf{X}_i\in\mathbb{R}^{M\times(d_i+1)}} \underbrace{\begin{bmatrix}c_{i,\varnothing}\\c_{ij_1}\\\vdots\\c_{ij_{d_i}}\end{bmatrix}}_{\mathbf{c}_i\in\mathbb{R}^{(d_i+1)}}.$$

To minimize the sum of squared deviations from the true coefficients, we use OLS, computing

$$\hat{\mathbf{c}}_i = (\mathbf{X}_i^\top\mathbf{X}_i)^{-1}\mathbf{X}_i^\top\mathbf{Y}_i.$$

We consider the limit of this estimate as $M \to \infty$. The consistency of the least squares estimator ensures that $\hat{\mathbf{c}}_i \xrightarrow{P} \mathbf{c}_i$. By the law of large numbers, we have

$$\mathbf{X}_i^\top\mathbf{X}_i = \sum_{m=1}^{M}\begin{bmatrix}1 & z_{j_1}^m & z_{j_2}^m & \dots & z_{j_{d_i}}^m\\z_{j_1}^m & (z_{j_1}^m)^2 & z_{j_1}^m z_{j_2}^m & \dots & z_{j_1}^m z_{j_{d_i}}^m\\z_{j_2}^m & z_{j_1}^m z_{j_2}^m & (z_{j_2}^m)^2 & \ddots & z_{j_2}^m z_{j_{d_i}}^m\\\vdots & \ddots & \ddots & \ddots & \vdots\\z_{j_{d_i}}^m & z_{j_1}^m z_{j_{d_i}}^m & z_{j_2}^m z_{j_{d_i}}^m & \dots & (z_{j_{d_i}}^m)^2\end{bmatrix} \xrightarrow{a.s.} M\cdot\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top].$$

Similarly, the vector $\mathbf{X}_i^\top\mathbf{Y}_i \xrightarrow{a.s.} M\cdot\mathbb{E}[\tilde{\mathbf{z}}_i\cdot Y_i(\mathbf{z})]$. Assuming the invertibility of the matrix $\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]$ (we show this explicitly for Bernoulli design below), we obtain in the limit,

$$\mathbf{c}_i = (M\cdot\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top])^{-1}M\cdot\mathbb{E}[\tilde{\mathbf{z}}_i\cdot Y_i(\mathbf{z})] = \mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]^{-1}\mathbb{E}[\tilde{\mathbf{z}}_i\cdot Y_i(\mathbf{z})].$$

Now, we return from our thought experiment to the actual experimental setting wherein a single instantiation $(\tilde{\mathbf{z}}_i, Y_i)$ is realized. We consider the estimator

$$\widehat{\mathbf{c}}_i \coloneqq \mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]^{-1}\,\tilde{\mathbf{z}}_i\,Y_i. \tag{4.2}$$

Note that the matrix $\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]^{-1}$ depends only on our experimental design and thus can be utilized by our estimator. The unbiasedness of this estimator follows from our aforementioned computation; by applying linearity,

$$\mathbb{E}[\widehat{\mathbf{c}}_i] = \mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]^{-1}\mathbb{E}[\tilde{\mathbf{z}}_i\,Y_i] = \mathbf{c}_i.$$

By applying linearity once more, we may obtain the unbiased estimator for the TTE,

$$\widehat{\mathrm{TTE}} = \frac{1}{n}\sum_{i=1}^{n}\langle(\mathbf{1}_{|\mathcal{S}_i^\beta|}-\mathbf{e}_1),\,\widehat{\mathbf{c}}_i\rangle = \frac{1}{n}\sum_{i=1}^{n}Y_i\langle\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]^{-1}(\mathbf{1}_{|\mathcal{S}_i^\beta|}-\mathbf{e}_1),\,\tilde{\mathbf{z}}_i\rangle. \tag{4.3}$$

For the specific setting of nonuniform Bernoulli design with $\beta = 1$, one can use the fact that $\mathbb{E}[z_{j_k}] = \mathbb{E}[z_{j_k}^2] = p_{j_k}$ and $\mathbb{E}[z_{j_k}z_{j_{k'}}] = p_{j_k}p_{j_{k'}}$ for $k \neq k'$ to compute,

$$\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top] = \begin{bmatrix}1 & p_{j_1} & p_{j_2} & \dots & p_{j_{d_i}}\\p_{j_1} & p_{j_1} & p_{j_1}p_{j_2} & \dots & p_{j_1}p_{j_{d_i}}\\p_{j_2} & p_{j_1}p_{j_2} & p_{j_2} & \ddots & p_{j_2}p_{j_{d_i}}\\\vdots & \ddots & \ddots & \ddots & \vdots\\p_{j_{d_i}} & p_{j_1}p_{j_{d_i}} & p_{j_2}p_{j_{d_i}} & \dots & p_{j_{d_i}}\end{bmatrix},$$

which we can invert to obtain

$$\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]^{-1} = \begin{bmatrix} 1+\sum_k \frac{p_{j_k}}{1-p_{j_k}} & -\frac{1}{1-p_{j_1}} & \cdots & \cdots & -\frac{1}{1-p_{j_{d_i}}} \\ -\frac{1}{1-p_{j_1}} & \frac{1}{p_{j_1}(1-p_{j_1})} & 0 & \cdots & 0 \\ \vdots & 0 & \frac{1}{p_{j_2}(1-p_{j_2})} & 0 & \vdots \\ \vdots & \vdots & 0 & \ddots & 0 \\ -\frac{1}{1-p_{j_{d_i}}} & 0 & \cdots & 0 & \frac{1}{p_{j_{d_i}}(1-p_{j_{d_i}})} \end{bmatrix}.$$

We calculate

$$\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]^{-1}\left(\mathbf{1}_{|\mathcal{S}_i^\beta|} - \mathbf{e}_1\right) = \begin{bmatrix} -\sum_{j\in\mathcal{N}_i}\frac{1}{1-p_j} & \frac{1}{p_{j_1}(1-p_{j_1})} & \cdots & \frac{1}{p_{j_{d_i}}(1-p_{j_{d_i}})} \end{bmatrix}^\top.$$

By plugging into equation (4.3), we obtain the explicit form for our estimator:

$$\widehat{\mathrm{TTE}}_{\mathrm{SNIPE}(1)} \coloneqq \frac{1}{n}\sum_{i=1}^n Y_i \sum_{j\in\mathcal{N}_i}\frac{z_j - p_j}{p_j(1-p_j)},$$

where SNIPE(1) refers to the fact that this is the SNIPE estimator with $\beta = 1$.

4.2 SNIPE in the general polynomial setting

For larger β, we can use the same least squares approach to obtain an unbiased estimate of the coefficient vector $\mathbf{c}_i$, and then plug this into equation (4.1) to obtain $\widehat{\mathrm{TTE}}$. The outcomes $Y_i(\mathbf{z})$ remain linear functions in $\tilde{\mathbf{z}}_i$ – albeit for a significantly longer $\tilde{\mathbf{z}}_i$ containing $|\mathcal{S}_i^\beta| = \sum_{k=0}^{\min(\beta,|\mathcal{N}_i|)}\binom{|\mathcal{N}_i|}{k}$ entries – so the same convergence results apply. Suppose we index the entries of $\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]$ by the sets $\mathcal{S}$ and $\mathcal{T}$ corresponding to the matrix row and column. By the independence and marginal treatment probabilities of nonuniform Bernoulli design, we see that

$$(\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top])_{\mathcal{S},\mathcal{T}} = \prod_{j\in\mathcal{S}\cup\mathcal{T}} p_j.$$

Working with this matrix is significantly more tedious, so we relegate the details to Appendix A. There, we show that $\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]$ is invertible by giving an explicit formula for the entries of its inverse. Then, we plug these entries into equation (4.3) to derive the following explicit formula for the estimator:

$$\widehat{\mathrm{TTE}}_{\mathrm{SNIPE}(\beta)} \coloneqq \frac{1}{n}\sum_{i=1}^n Y_i \sum_{\mathcal{S}\in\mathcal{S}_i^\beta} g(\mathcal{S}) \prod_{j\in\mathcal{S}}\frac{z_j - p_j}{p_j(1-p_j)}, \tag{4.4}$$

where we define the coefficient function $g : 2^{[n]} \to \mathbb{R}$ such that

$$g(\mathcal{S}) = \prod_{s\in\mathcal{S}}(1-p_s) - \prod_{s\in\mathcal{S}}(-p_s)$$

for each $\mathcal{S} \subseteq [n]$ and $g(\varnothing) = 0$. When it is clear from context that $\widehat{\mathrm{TTE}}$ refers to the SNIPE estimator, we suppress the subscript for cleaner notation.

Remark 5. We pause here to again emphasize that our technique gives us an unbiased estimate of the *coefficient vector* $\mathbf{c}_i$ for each individual i. From here, we leverage the linearity of expectation to obtain an unbiased estimator for the TTE, which is a linear function of these $\mathbf{c}_i$ coefficients. This same strategy can be

applied to develop estimators for any causal effect that is linear in the $\mathbf{c}_{i,\mathcal{S}}$ coefficients. We highlight some other potential estimands and the explicit form of their estimators in Appendix E. The techniques that we discuss in Section 5 can be used to establish further properties of these estimators.

This estimator can be evaluated in $O(nd_{\text{in}}^{\beta})$ time and only utilizes structural information about the graph (not any influence coefficients $c_{i,\mathcal{S}}$). Structurally, the estimator takes the form of a weighted average of the outcomes Y_i of each individual i, where the weights themselves are functions of the treatment assignments of all members j of the in-neighborhood $\mathcal{N}_i$. To make use of the low-order interference assumption, the estimator separately scales the effect of treatment of each sufficiently small subset of $\mathcal{N}_i$ using the scaling function $g(\mathcal{S})$. The definition of this $g(\mathcal{S})$ ensures the unbiasedness of the estimator.

In the special case of a uniform treatment probability $p_i = p$ across all nodes, we can simplify this estimator to show that it is only a function of the number of treated individuals in i's neighborhood and not the identities. Let $\mathcal{T} = \{j : z_j = 1\}$ denote the set of treated units. We can rewrite the estimator as follows:

$$\begin{aligned}
\widehat{\text{TTE}} &= \frac{1}{n}\sum_{i=1}^{n} Y_i \sum_{\mathcal{S}\in\mathcal{S}_i^{\beta}} \left(\prod_{j\in\mathcal{S}} \frac{z_j - p}{p} - \prod_{j\in\mathcal{S}} \frac{z_j - p}{p-1} \right) \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i \sum_{k=0}^{\min(\beta,|\mathcal{N}_i\cap\mathcal{T}|)} \sum_{\substack{\mathcal{A}\subseteq\mathcal{N}_i\cap\mathcal{T} \\ |\mathcal{A}|=k}} \sum_{\substack{\mathcal{B}\subseteq\mathcal{N}_i\setminus\mathcal{T} \\ |\mathcal{B}|\le\beta-k}} \left(\left(\frac{1-p}{p}\right)^k (-1)^{|\mathcal{B}|} - (-1)^k \left(\frac{p}{1-p}\right)^{|\mathcal{B}|} \right) \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i \sum_{k=0}^{\min(\beta,|\mathcal{N}_i\cap\mathcal{T}|)} \binom{|\mathcal{N}_i\cap\mathcal{T}|}{k} \sum_{\ell=0}^{\min(\beta-k,|\mathcal{N}_i\setminus\mathcal{T}|)} \binom{|\mathcal{N}_i\setminus\mathcal{T}|}{\ell} (-1)^{k+\ell} \left(\left(\frac{p-1}{p}\right)^k - \left(\frac{-p}{1-p}\right)^{\ell} \right).
\end{aligned} \tag{4.5}$$

Thus, while our estimation guarantees allow for heterogeneity in the potential outcomes model, when treatment probabilities are uniform, computing the estimator does not depend on the identity of the treated individuals, but only the number of treated neighbors. As a result, the estimator can be evaluated in $O(n\beta^2)$ time, which is a significant improvement compared to the $O(nd_{\text{in}}^{\beta})$ computational complexity when the treatment probabilities are nonuniform.

4.3 Connection to other estimator classes

Our estimator takes the form of a linear weighted estimator

$$\widehat{\text{TTE}} = \frac{1}{n}\sum_{i=1}^{n} Y_i \cdot w_i(\mathbf{z}),$$

under specifically constructed weight functions $w_i : \{0,1\}^n \to \mathbb{R}$. From this perspective, we can draw connections between our estimator and others appearing in the literature.

4.3.1 Horvitz–Thompson estimator

First, we show that in the special case where $|\mathcal{N}_i| \le \beta$, our estimator is identical to the classical Horvitz–Thompson estimator. In this case, the restriction $|\mathcal{S}| \le \beta$ is satisfied for every $\mathcal{S} \subseteq \mathcal{N}_i$, so that $\mathcal{S}_i^{\beta}$ includes all subsets of $\mathcal{N}_i$. By using this, we may simplify

$$
\begin{aligned}
w_i(\mathbf{z}) &= \sum_{\mathcal{S}\subseteq\mathcal{N}_i} g(\mathcal{S}) \prod_{j\in\mathcal{S}} \frac{z_j - p_j}{p_j(1-p_j)} \\
&= \sum_{\mathcal{S}\subseteq\mathcal{N}_i} \prod_{j\in\mathcal{S}} \frac{z_j - p_j}{p_j} - \sum_{\mathcal{S}\subseteq\mathcal{N}_i} \prod_{j\in\mathcal{S}} \frac{-(z_j - p_j)}{(1-p_j)} && \text{(by definition of } g(\mathcal{S})) \\
&= \prod_{j\in\mathcal{N}_i} \left(1 + \frac{z_j - p_j}{p_j}\right) - \prod_{j\in\mathcal{N}_i} \left(1 - \frac{z_j - p_j}{1-p_j}\right) && \text{(distributivity)} \\
&= \prod_{j\in\mathcal{N}_i} \frac{z_j}{p_j} - \prod_{j\in\mathcal{N}_i} \frac{1-z_j}{1-p_j} \\
&= \frac{\mathbb{1}(\mathbf{z} \text{ treats all of } \mathcal{N}_i)}{\Pr(\mathbf{z} \text{ treats all of } \mathcal{N}_i)} - \frac{\mathbb{1}(\mathbf{z} \text{ treats none of } \mathcal{N}_i)}{\Pr(\mathbf{z} \text{ treats none of } \mathcal{N}_i)}.
\end{aligned}
$$

Thus,

$$
\widehat{\mathrm{TTE}}_{\mathrm{SNIPE}} = \frac{1}{n}\sum_{i=1}^{n} Y_i \left(\frac{\mathbb{1}(\mathbf{z} \text{ treats all of } \mathcal{N}_i)}{\Pr(\mathbf{z} \text{ treats all of } \mathcal{N}_i)} - \frac{\mathbb{1}(\mathbf{z} \text{ treats none of } \mathcal{N}_i)}{\Pr(\mathbf{z} \text{ treats none of } \mathcal{N}_i)} \right) = \widehat{\mathrm{TTE}}_{HT},
$$

so we exactly recover the Horvitz–Thompson estimator. As a result, when β is sufficiently large relative to the degree of the nodes in the graph, our estimator is very similar to the Horvitz–Thompson estimator, only differing for the nodes that have graph degrees larger than β. In this sense, under Bernoulli randomization, our estimator can be thought of as a generalization of Horvitz–Thompson to additionally account for low polynomial degree structure, which is most relevant for simplifying the potential outcomes associated with high-degree vertices.

4.3.2 Pseudoinverse estimator

The key technical steps of deriving our estimator can be described as constructing unbiased estimators for each unit i's contribution to the TTE using a connection to OLS for linear models, and subsequently averaging the unbiased estimates due to the fact that the causal estimand is a linear function of these individual contributions. This overall technique has also appeared in the previous literature in semiparametric estimation, with one clear example described by Swaminathan et al. [45] in a seemingly different context of off-policy evaluation for online recommendation systems. In their model, a context $x \in X$ arrives, at which point the principal selects a tuple (or *slate*) of actions $\mathbf{s} = (s_1, \ldots, s_\ell)$ and observes a random reward r based on the interaction between the context and the slate. The authors make a linearity assumption that posits that $V(x, \mathbf{s}) = \mathbb{E}_r[r|x, \mathbf{s}] = \mathbf{1}_s^\top \phi_x$, where $\mathbf{1}_s^\top$ indicates the choice of a particular action in each entry of the tuple, and θ_x is a context-specific reward weight vector associated to context x. In our setting, the context x is an individual i, the reward weights θ_x are their effect coefficients $\mathbf{c}_i$, and the slate indicator vector $\mathbf{1}_s^\top$ is our treated subsets vector $\tilde{\mathbf{z}}_i$.

A primary goal of ref. [45] is to estimate ϕ_x, which can be used to inform a good slate selection policy. The mean squared error of an estimate $\mathbf{w}$ for r is $\mathbb{E}_\mu[(\mathbf{1}_s^\top \mathbf{w} - r)^2]$, where μ encodes the random selection of the context, slate, and reward. In this framing, the least squares estimator

$$
\bar{\theta}_x = (\mathbb{E}_\mu[\mathbf{1}_s \mathbf{1}_s^\top | x])^\dagger \mathbb{E}_\mu[r \mathbf{1}_s^\top | x]
$$

is the minimizer of the mean squared error (MSE) with the minimum norm. As the reward distribution is unknown, it is replaced with an empirical estimate of r given past data. We can perform the substitutions as described in the previous paragraph, noting that Bernoulli design ensures that $\mathbb{E}[\tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i^\top]$ is invertible (Appendix A), to recover our estimator $\widehat{\mathbf{c}}_i$ from (4.2).

4.3.3 Riesz estimator

In their recent work, Harshaw et al. [46] consider a very general causal inference framework wherein a treatment $\mathbf{z}$ is drawn from an underlying experimental design distribution over an intervention set $\mathcal{Z}$. We

observe an outcome $Y_i(\mathbf{z})$ for each unit $i \in [n]$, where we assume that Y_i belongs to a *model space* $\mathcal{M}_i$, which is a subspace of $\mathcal{Y}$ containing all measurable and square-integrable (with respect to the distribution over $\mathcal{Z}$) functions $\mathcal{Z} \to \mathbb{R}$. Each individual is additionally endowed with a linear effect functional $\theta_i : \mathcal{Y} \to \mathbb{R}$, and the goal is to estimate the *average individual treatment effect* $\tau = \frac{1}{n}\sum_{i=1}^{n}\theta_i(Y_i)$.

In our setting, $\mathcal{Z} = \{0,1\}^n$ with the nonuniform Bernoulli design distribution (i.e., $\Pr(z_i = 1) = p_i$, with $\{z_i\}$ mutually independent). $\mathcal{Y}$ consists of all linear functions over the basis $\{\prod_{j\in\mathcal{S}} z_j : \mathcal{S} \subseteq [n], |\mathcal{S}| \le \beta\}$, and each $\mathcal{M}_i$ restricts to the linear functions over $\{\prod_{j\in\mathcal{S}} z_j : \mathcal{S} \in \mathcal{S}_i^{\beta}\}$. Finally, each unit i has effect functional $\theta_i(Y_i) = Y_i(\mathbf{1}) - Y_i(\mathbf{0})$, so that $\tau = \mathrm{TTE}$.

Under two assumptions – correct model specification (needed in our work as well) and positivity (always satisfied in our setting since Bernoulli design ensures each treatment in $\mathcal{Z}$ occurs with positive probability) – the Riesz representation theorem guarantees the existence of an unbiased estimator for τ, referred to as a Riesz estimator, of the form

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z})\psi_i(\mathbf{z}),$$

where $\psi_i \in \mathcal{M}_i$ is a Riesz representor for θ_i with the property that

$$\theta_i(Y) = \mathbb{E}_{\mathbf{z}}[\psi_i(\mathbf{z})Y(\mathbf{z})] \qquad (\star)$$

for each $Y \in \mathcal{M}_i$. As each model space $\mathcal{M}_i$ has dimension $|\mathcal{S}_i^{\beta}|$, we can identify ψ_i by solving the linear system that verifies that $(\star)$ holds for each function in some basis for $\mathcal{M}_i$. A canonical choice is the standard basis, giving rise to equations

$$\theta_i\left(\prod_{j'\in\mathcal{S}'} z_{j'}\right) = \mathbb{I}(\mathcal{S}' \neq \varnothing) = \mathbb{E}_{\mathbf{z}}\left[\psi_i(\mathbf{z})\prod_{j'\in\mathcal{S}'} z_{j'}\right],$$

for each $\mathcal{S}' \in \mathcal{S}_i^{\beta}$. The choice $\psi_i(\mathbf{z}) = \sum_{\mathcal{S}\in\mathcal{S}_i^{\beta}} g(\mathcal{S})\prod_{j\in\mathcal{S}} \frac{z_j - p_j}{p_j(1-p_j)}$ is a solution to this system (Appendix B) and gives rise to our estimator.

5 Properties of estimator under Bernoulli design

The following theorem summarizes the key properties of our estimator.

Theorem 1. *Under a potential outcomes model satisfying the neighborhood interference assumption with polynomial degree at most β, the estimator defined in* (4.4) *is unbiased with variance upper bounded by*

$$\frac{d_{\mathrm{in}}\, d_{\mathrm{out}}\, {Y_{\max}}^2}{n} \cdot \left(\frac{e d_{\mathrm{in}}}{\beta} \cdot \max\left(4\beta^2, \frac{1}{p(1-p)}\right)\right)^{\beta},$$

where each $p_i \in [p, 1-p]$ and $p > 0$.

Notably, a sequence of networks with $n \to \infty$ and $d = o(\log n)$ has variance asymptotically approaching 0. We defer the proof of this theorem to Appendix B. Rather than appealing to the convergence properties of the pseudoinverse estimator to establish unbiasedness, we present an alternate combinatorial proof. To bound the variance, we carefully consider different possible overlapping subsets of individuals to separately bound many covariance terms that make up the overall variance expression.

To understand the variance bounds for our estimator, we can compare against the variance of Horvitz–Thompson under a Bernoulli design. In the simple setting of a d-regular graph and uniform Bernoulli(p) randomization, ref. [17] showed that the Horvitz–Thompson estimator has a variance that is lower bounded by $\Omega(1/np^d)$. In contrast, the variance of our estimator only scales polynomially in the degree d, but exponentially in the polynomial degree β, which is achieved by simply changing the estimator, without

requiring any additional clustering structure of the graph and without utilizing complex RDs. This is a significant gain when the polynomial degree β is significantly lower than the graph degree d. The simplest setting of $\beta = 1$ already expresses all potential outcomes models that satisfy additivity of main effects and additivity of interference, as defined in ref. [18]; this subsumes all linear models that are commonly used in the practical literature, yet which require additional homogeneity assumptions.

We also remark that when $\beta = d$, both the Horvitz–Thompson estimator and our estimator are unbiased, for the same class of functions. When $\beta < d$, the Horvitz–Thompson estimator is unbiased for a strictly larger class of functions than our estimator, precisely those characterized by $\beta = d$. In this way, if the practitioner can use domain knowledge to argue that the true potential outcomes model belongs to the class of functions parametrized by β from equation (3.2) for $\beta < d$, then our estimator provides an advantage over the Hortvitz-Thompson estimator. This is because both estimators are unbiased but the variance of our estimator does not have exponential dependence on the graph degree. However, depending on the flexibility that the practitioner desires or needs to express in the potential outcomes, there is no clear winner. For example, if one desires a fully nonparametric potential outcomes to capture the most general neighborhood interference settings, they might set $\beta = d$. Then, both our estimator and the Horvitz–Thompson estimator are unbiased, and both have variance scaling exponentially in d. On the other hand, suppose anonymous interference was satisfied for a particular application and the potential outcomes could be modeled:

$$Y_i(\mathbf{z}) = c_0 + c_1 z_i + c_2 \left(\sum_{j \in \mathcal{N}_i} z_j \right).$$

Then, using an OLS estimator would give an unbiased estimate with lower variance than our estimator. Thus, our model and estimator can be viewed as simply "adding to the toolbox" that practitioners may use depending on how expressive they need their potential outcomes models to be.

In the special setting of uniform Bernoulli design and $\beta = 1$, our estimator as stated in (4.5) is the same as an estimator presented in ref. [22]. They consider a fully nonparametric setting under anonymous interference, in which the goal is to estimate the derivative of the total outcomes under changes of the population-wide treatment probability. As the derivative can be estimated by locally linearizing the outcomes function, they derive the special case of the estimator in (4.5) under $\beta = 1$ by taking the derivative of the expected population outcomes under a Bernoulli randomization, constructed by inverse propensity weights. This suggests that under a fully nonparametric setting, our estimator may be used for estimating an appropriately defined local estimand. There may also be opportunities to perform variance reduction on our estimator given knowledge of the graph structure, as proposed in ref. [22]; however, their solution requires anonymous interference, and it is not clear how to extend their solution concept beyond $\beta = 1$, anonymous interference, and uniform treatment probabilities.

5.1 Minimax lower bound on mean squared error

To understand the optimality of our estimator, we construct a lower bound on the minimax optimal mean squared error rate. In particular, we show that for a setting with a d-regular graph and sufficiently-small uniform treatment probabilities p, the best achievable mean squared error is lower bounded by $\Omega\left(\frac{1}{np^\beta}\right)$.

Theorem 2. (Minimax lower bound) *For any n, d, β, p with $p^\beta < 0.16$, and any estimator $\widehat{\mathrm{TTE}}$, there exists a causal network on n nodes with maximum degree d and effect coefficients $\{c_{i,S}\}_{i \in [n], S \in \mathcal{S}_i}$ for which the minimax squared error under uniform treatment probabilities $p_i = p$ is bounded below by*

$$\mathbb{E}[(\widehat{\mathrm{TTE}} - \mathrm{TTE})^2] = \Omega\left(\frac{1}{np^\beta}\right).$$

For a β-order interactions model, estimating the TTE requires being able to measure the network effect of the size β subsets of a unit's neighborhood. The probability that a set of size β is jointly assigned to treatment is p^β. As a result, the scaling of $\frac{1}{p^\beta}$ is somewhat intuitive.

The proof of Theorem 2 (given in Appendix C) uses a generalized variation of LeCam's method for fuzzy hypothesis testing [47], [48, Sec. 2.7.4]. We reduce the problem of TTE estimation to the creation of a hypothesis test to distinguish between two priors. We consider a setting with a d-regular graph, uniform treatment probabilities $p_i = p$, and two Gaussian priors Γ_0, Γ_1 over the effect coefficients. The priors have the same variances but with shifted means such that $|\mathbb{E}_{\Gamma_0}[\mathrm{TTE}] - \mathbb{E}_{\Gamma_1}[\mathrm{TTE}]| = 2\delta$ for some carefully tuned parameter δ. The failure of hypothesis tests that rely on $\widehat{\mathrm{TTE}}$ is attributable to one of two factors: (1) a significant shift in TTE brought about by the variability of the coefficients, or (2) inaccuracy of the estimator. We bound the probability of the former with a Gaussian tail bound. We bound the latter in terms of the Kullback-Leibler (KL)-divergence between the distributions over $(\mathbf{Y}, \mathbf{z})$ induced by Γ_0 and Γ_1. The rest of the argument involves a calculation of this KL-divergence and a selection of δ to ensure a nonvanishing error probability.

While our lower bound clearly indicates that the exponential dependence on β as exhibited by $p^{-\beta}$ is necessary, a notable difference between our lower and upper bounds is the dependence on the graph degree d. Recall that the upper bound on the variance of our SNIPE estimator in Section 1 scales roughly as $d^{\beta+2}$; however, our lower bound result has no dependence on d. We attribute this gap to a weakness in the analyses and believe that it can be tightened with more careful calculation. In the upper bound, the d^β arises as a bound on the sum of binomial coefficients of the form $\binom{d}{k}$ for $k \le \beta$. While this bound is precise asymptotically for a regime, where β is a small constant and d is large, it is loose in the most general case where $\beta = d$, where we know 2^d is sufficient. On the other hand, our lower bound argument considers a setting with a d-regular graph, but it only uses the degree of the graph to normalize the parameters on the prior distributions, such that we see no dependence on d in the final lower bound.

5.2 Central limit theorem

In this section, we show $\widehat{\mathrm{TTE}}_{\mathrm{SNIPE}}$ is asymptotically normal using Stein's method, a common approach to proving central limit theorems [3,49–52]. In particular, we use Theorem 3.6 from Ross [53], which we restate below for convenience.

[53, Theorem 3.6]. Let $X_1, \ldots, X_n$ be random variables such that $\mathbb{E}[X_i^4] < \infty$, $\mathbb{E}[X_i] = 0$, $v^2 = \mathrm{Var}(\sum_i X_i)$, and define $W = \sum_i X_i / v$. Let collection $(X_1, \ldots, X_n)$ have dependency neighborhoods D_i for $i \in [n]$, and also define $D \coloneqq \max_{1 \le i \le n} |D_i|$. Then, for Z a standard normal random variable,

$$d_{\mathrm{W}}(W, Z) \le \frac{D^2}{v^3} \sum_{i=1}^{n} \mathbb{E}|X_i|^3 + \frac{\sqrt{28}\, D^{3/2}}{\sqrt{\pi}\, v^2} \sqrt{\sum_{i=1}^{n} \mathbb{E}[X_i^4]}, \tag{5.1}$$

where $d_{\mathrm{W}}(W, Z)$ is the Wasserstein distance between W and Z.

Our estimator can be written in the form of $\widehat{\mathrm{TTE}} = \frac{1}{n} \sum_{i \in [n]} Y_i \, w_i(\mathbf{z})$, where Y_i and $w_i(\mathbf{z})$ depend only on the treatment assignments of individuals $j \in \mathcal{N}_i$. We apply Theorem 3.6 from ref. [53], to the random variables $X_i \coloneqq \frac{1}{n}(Y_i w_i(\mathbf{z}) - \mathbb{E}[Y_i w_i(\mathbf{z})])$, such that by construction $W = \sum_i X_i / v = (\widehat{\mathrm{TTE}} - \mathrm{TTE})/v$ and $\widehat{\mathrm{TTE}} = vW + \mathrm{TTE}$. An application of Theorem 3.6 from ref. [53] implies that $\widehat{\mathrm{TTE}}$ is asymptotically normal as long as the bound in (5.1) converges to 0 as n approaches infinity. As the size of the dependency neighborhoods D, the moments of X_i, and the scaling of the variance v all depend on the network parameters, for simplicity, we additionally assume in the conditions of Theorem 3 that d, $Y_{\max}$, and β are bounded above by a constant, and we also assume the treatment probability p is also lower bounded as a function of n.

Assumption 3. For some constant $c > 0$, we have $n \cdot \mathrm{Var}[\widehat{\mathrm{TTE}}] \to c$ as $n \to \infty$.

This is a typical assumption made in the literature [3,51,54] that rules out degenerate cases, such as all the potential outcomes being 0, that would result in the estimator having an unnaturally low variance scaling smaller than $1/n$.

Theorem 3. (Central limit theorem) *Under Assumptions* 1–3, *and assuming that* d, $Y_{\max}$, *and* β *are all* $O(1)$ *with respect to* n, $p = \omega(n^{-1/4\beta})$, *and* $p_i \le \frac{1}{2}$ *for all* i, *the normalized error* $(\widehat{\mathrm{TTE}} - \mathrm{TTE})/\nu$ *converges in distribution to a standard normal random variable as* $n \to \infty$, *for* $\nu^2 = \mathrm{Var}[\widehat{\mathrm{TTE}}]$.

Theorem 3 states that our estimator is asymptotically normal, assuming the boundedness of the magnitude of the potential outcomes, the polynomial degree, and the network degree. For the complete proof, refer to Appendix D. The proof follows from a straightforward application of Theorem 3.6 from ref. [53] with appropriate bounds for the moments $\mathbb{E}|X_i|^3$ and $\mathbb{E}[X_i^4]$, the dependency neighborhood size D, and the variance ν. The dependency neighborhood for variable X_i is the set $\{j \in [n] \text{ s.t. } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset\}$, i.e., the set of individuals j that share in-neighbors with individual i. This follows from the observation that both X_i and X_j are a function of the treatment variables z_k for shared in-neighbors $k \in \mathcal{N}_i \cap \mathcal{N}_j$. The size of the largest dependency neighborhood is thus bounded by $D \le d_{\text{in}} d_{\text{out}}$. The moments $\mathbb{E}|X_i|^3$, $\mathbb{E}[X_i^4]$ will be bounded as a function of $Y_{\max}$ as defined in (3.3) along with the network degree d, treatment probability p, and the polynomial degree β. The boundedness assumptions are used to argue that these moments do not grow too quickly in n, so that the Wasserstein distance in (5.1) converges to 0. The details of these calculations are deferred to Appendix D. We remark that the strict boundedness assumption can be relaxed so that d, $Y_{\max}$, and β are polylogarithmic with respect to n, so long as we tighten the lower bound on the growth of p by a corresponding logarithmic factor.

5.3 Variance estimator

The central limit theorem result in Theorem 3 implies that we can construct asymptotically valid confidence intervals by $[\widehat{\mathrm{TTE}} - \nu\Phi^{-1}(1-\alpha), \widehat{\mathrm{TTE}} + \nu\Phi^{-1}(1-\alpha)]$ if we knew ν, where Φ indicates the cdf of a standard normal. Since the practitioner typically does not know ν, we construct a conservative estimator for ν by applying the approach from Aronow and Samii [3,55]. We begin by rewriting the SNIPE estimator as a sum over possible exposures $\mathbf{x}$:

$$\widehat{\mathrm{TTE}} = \frac{1}{n}\sum_{i\in[n]} Y_i(\mathbf{z}) w_i(\mathbf{z}) = \frac{1}{n}\sum_{i\in[n]}\sum_{\mathbf{x}\in\{0,1\}^{|\mathcal{N}_i|}} \mathbb{I}(\mathbf{z}_{\mathcal{N}_i} = \mathbf{x}) Y_i(\mathbf{x}) w_i(\mathbf{x}).$$

Note that both $Y_i(\mathbf{x})$ and $w_i(\mathbf{x})$ are deterministic, and the only randomness is expressed in the indicator function $\mathbb{I}(\mathbf{z}_{\mathcal{N}_i} = \mathbf{x})$. Overloading notation, we let Y_i and w_i be expressed only as a function of $\mathbf{z}_{\mathcal{N}_i}$, where we assume the order β is clearly specified. Then, the variance of our estimator is given by

$$\mathrm{Var}[\widehat{\mathrm{TTE}}] = \frac{1}{n^2}\sum_{i,j\in[n]}\sum_{\mathbf{x}\in\{0,1\}^{|\mathcal{N}_i|}}\sum_{\mathbf{x}'\in\{0,1\}^{|\mathcal{N}_j|}} Y_i(\mathbf{x}) w_i(\mathbf{x}) \quad Y_j(\mathbf{x}') w_j(\mathbf{x}') \quad \mathrm{Cov}(\mathbb{I}(\mathbf{z}_{\mathcal{N}_i} = \mathbf{x}), \mathbb{I}(\mathbf{z}_{\mathcal{N}_j} = \mathbf{x}')).$$

If $\mathbb{P}(\{\mathbf{z}_{\mathcal{N}_i} = \mathbf{x}\} \cap \{\mathbf{z}_{\mathcal{N}_j} = \mathbf{x}'\}) > 0$, an unbiased estimate for the corresponding term in the variance expression is given by

$$\frac{\mathbb{I}(\mathbf{z}_{\mathcal{N}_i} = \mathbf{x})\mathbb{I}(\mathbf{z}_{\mathcal{N}_j} = \mathbf{x}')}{\mathbb{P}(\{\mathbf{z}_{\mathcal{N}_i} = \mathbf{x}\} \cap \{\mathbf{z}_{\mathcal{N}_j} = \mathbf{x}'\})} Y_i(\mathbf{x}) w_i(\mathbf{x}) Y_j(\mathbf{x}') w_j(\mathbf{x}') \mathrm{Cov}(\mathbb{I}(\mathbf{z}_{\mathcal{N}_i} = \mathbf{x}), \mathbb{I}(\mathbf{z}_{\mathcal{N}_j} = \mathbf{x}')).$$

Otherwise, if $\mathbb{P}(\{\mathbf{z}_{\mathcal{N}_i} = \mathbf{x}\} \cap \{\mathbf{z}_{\mathcal{N}_j} = \mathbf{x}'\}) = 0$, by the same technique as given in ref. [55], using the observation that $\mathrm{Cov}(\mathbb{I}(\mathbf{z}_{\mathcal{N}_i} = \mathbf{x}), \mathbb{I}(\mathbf{z}_{\mathcal{N}_j} = \mathbf{x}')) = \mathbb{P}(\mathbf{z}_{\mathcal{N}_i} = \mathbf{x})\mathbb{P}(\mathbf{z}_{\mathcal{N}_j} = \mathbf{x}')$, the corresponding term in the variance expression can be approximated by the inflated estimate

$$\frac{1}{2}(\mathbb{I}(\mathbf{z}_{\mathcal{N}_i} = \mathbf{x})\mathbb{P}(\mathbf{z}_{\mathcal{N}_i} = \mathbf{x}) Y_i^2(\mathbf{x}) w_i^2(\mathbf{x}) + \mathbb{I}(\mathbf{z}_{\mathcal{N}_j} = \mathbf{x}')\mathbb{P}(\mathbf{z}_{\mathcal{N}_j} = \mathbf{x}') Y_j^2(\mathbf{x}') w_j^2(\mathbf{x}')).$$

The expectation of this estimate will be higher than the true term in the variance via Cauchy–Schwarz inequality. Therefore, a conservative variance estimator $\widehat{\mathrm{Var}}(\widehat{\mathrm{TTE}}_{\mathrm{SNIPE}})$ is given by

$$\frac{1}{n^2}\sum_i \sum_{j\in\mathcal{M}_i}\sum_{\mathbf{x}\in\{0,1\}^{|\mathcal{N}_i\cup\mathcal{N}_j|}} \frac{\mathbb{I}(\mathbf{z}_{\mathcal{N}_i\cup\mathcal{N}_j}=\mathbf{x})}{\mathbb{P}(\mathbf{z}_{\mathcal{N}_i\cup\mathcal{N}_j}=\mathbf{x})} Y_i(\mathbf{x})w_i(\mathbf{x})Y_j(\mathbf{x})w_j(\mathbf{x})\ \mathrm{Cov}(\mathbb{I}(\mathbf{z}_{\mathcal{N}_i}=\mathbf{x}_{\mathcal{N}_i}),\mathbb{I}(\mathbf{z}_{\mathcal{N}_j}=\mathbf{x}_{\mathcal{N}_j}))$$
$$+\frac{1}{n^2}\sum_i\sum_{\mathbf{x}\in\{0,1\}^{|\mathcal{N}_i|}}\mathbb{I}(\mathbf{z}_{\mathcal{N}_i}=\mathbf{x})\mathbb{P}(\mathbf{z}_{\mathcal{N}_i}=\mathbf{x})Y_i^2(\mathbf{x})w_i^2(\mathbf{x})\sum_{j\in\mathcal{M}_i}(2^{|\mathcal{N}_j|}-2^{|\mathcal{N}_j\setminus\mathcal{N}_i|}),$$

where

$$\mathrm{Cov}(\mathbb{I}(\mathbf{z}_{\mathcal{N}_i}=\mathbf{x}_{\mathcal{N}_i}),\mathbb{I}(\mathbf{z}_{\mathcal{N}_j}=\mathbf{x}_{\mathcal{N}_j}))=\prod_{k\in\mathcal{N}_i\cup\mathcal{N}_j}p_k^{x_k}(1-p_k)^{1-x_k}\left(1-\prod_{k'\in\mathcal{N}_i\cap\mathcal{N}_j}p_{k'}^{x_{k'}}(1-p_{k'})^{1-x_{k'}}\right).$$

In the Section 6, we compare the empirical variance of the SNIPE estimator $\widehat{\mathrm{TTE}}$ with this conservative variance estimate $\widehat{\mathrm{Var}}(\widehat{\mathrm{TTE}})$ in simulations. We also compare against the worst-case variance bound from Theorem 1. Most notable is that both the conservative variance estimate and the worst-case variance bound are orders of magnitude larger than the empirical variance, suggesting that there is significant work needed to develop tighter variance estimates.

6 Experimental results

Using computational experiments on simulated data, we compare the performance of our estimator with existing estimators. By using an Erdös–Rényi model, we generate random directed graphs of n nodes for a population of n individuals. Figure 1 shows results from networks made using the Erdös–Rényi model with n nodes and probability $p_{\text{edge}} = 10/n$ of an edge existing between any two nodes. Hence, the expected in-degree and out-degree of each node is 10. For degree β, we construct the same potential outcomes model as in ref. [43]:

$$Y_i(\mathbf{z}) = c_{i,\varnothing} + \sum_{j\in\mathcal{N}_i}\tilde{c}_{ij}z_j + \sum_{\ell=2}^{\beta}\left(\frac{\sum_{j\in\mathcal{N}_i}\tilde{c}_{ij}z_j}{\sum_{j\in\mathcal{N}_i}\tilde{c}_{ij}}\right)^{\ell}, \tag{6.1}$$

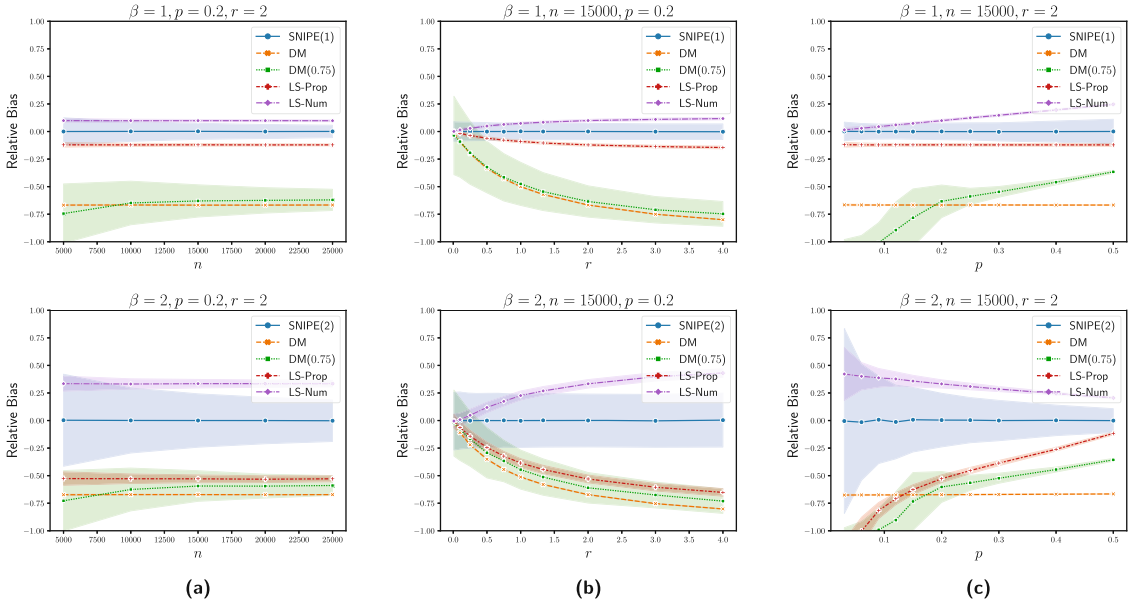


Figure 1: Plots visualizing the performance of various TTE estimators under Bernoulli design on Erdös–Rényi networks for both linear and quadratic potential outcomes models. The height of each line on a plot depicts the experimental relative bias of the estimator and the shaded width depicts the experimental standard deviation. The SNIPE estimator is parametrized by β, the degree of the potential outcomes model. (a) Varying population size, (b) varying direct:indirect effects, and (c) varying treatment budget.

where $c_{i,\varnothing} \sim U[0,1]$, $\tilde{c}_{ii} \sim U[0,1]$, and for $i \neq j$, $\tilde{c}_{ij} = v_j|\mathcal{N}_i|/\sum_{k:(k,j)\in E}|\mathcal{N}_k|$ for $v_j \sim U[0,r]$, where r denotes a hyperparameter that governs the magnitude of the network effects relative to the direct effects. We represent the magnitude of individual j's influence by the parameter v_j. This influence is shared among individual j's out-neighbors proportional to their in-degrees.

6.1 Other estimators

We compare the performance of SNIPE with the performance of least-squares regression and difference-in-means estimators, also as in ref. [43]. Although the Horvitz–Thompson estimator is unbiased in this setting, under unit Bernoulli RD its variance is very high in practice. Thus, we omit this estimator from our experimental results and comparison. Another related estimator is the Hájek estimator, which is only approximately unbiased but with lower variance than the Horvitz–Thompson estimator. However, under our RD, its variance is still very high in practice. In addition, as we implemented a Bernoulli RD on a graph with expected network degree of 10, both the Hájek and Horvitz–Thompson estimators consistently took a value of 0, giving us no meaningful results to consider. For these reasons, we chose to omit these two estimators from our experimental results and comparison. The simplest difference-in-means estimator is the difference between the average outcome of individuals assigned to treatment and the average outcome of individuals assigned to control, given by

$$\widehat{\mathrm{TTE}}_{\mathrm{DM}} = \frac{\sum_{i\in[n]} z_i Y_i}{\sum_{i\in[n]} z_i} - \frac{\sum_{i\in[n]}(1-z_i)Y_i}{\sum_{i\in[n]}(1-z_i)}. \tag{6.2}$$

This estimator does not take into account any information about each individual's neighborhood and is biased under network interference. We also consider a modified version of this estimator that uses information about the number of treated neighbors of each individual. Let U_i denote the number of individuals in $\mathcal{N}_i\backslash\{i\}$ assigned to treatment, and let $\tilde{U}_i$ denote the number of neighbors individuals in $\mathcal{N}_i\backslash\{i\}$ assigned to control. Then, the estimator is given by

$$\widehat{\mathrm{TTE}}_{\mathrm{DM}(\lambda)} = \frac{\sum_{i\in[n]} z_i \mathbb{I}(U_i \geq \lambda) Y_i}{\sum_{i\in[n]} z_i \mathbb{I}(U_i \geq \lambda)} - \frac{\sum_{i\in[n]}(1-z_i)\mathbb{I}(\tilde{U}_i \geq \lambda) Y_i}{\sum_{i\in[n]}(1-z_i)\mathbb{I}(\tilde{U}_i \geq \lambda)}, \tag{6.3}$$

for some user-defined tolerance $\lambda \in [0,1]$. We set $\lambda = 0.75$ for our experiments. Note that $\widehat{\mathrm{TTE}}_{\mathrm{DM}(\lambda)}$ counts an individual i's outcome only when at least λ of their neighborhood is assigned to the same treatment as them.

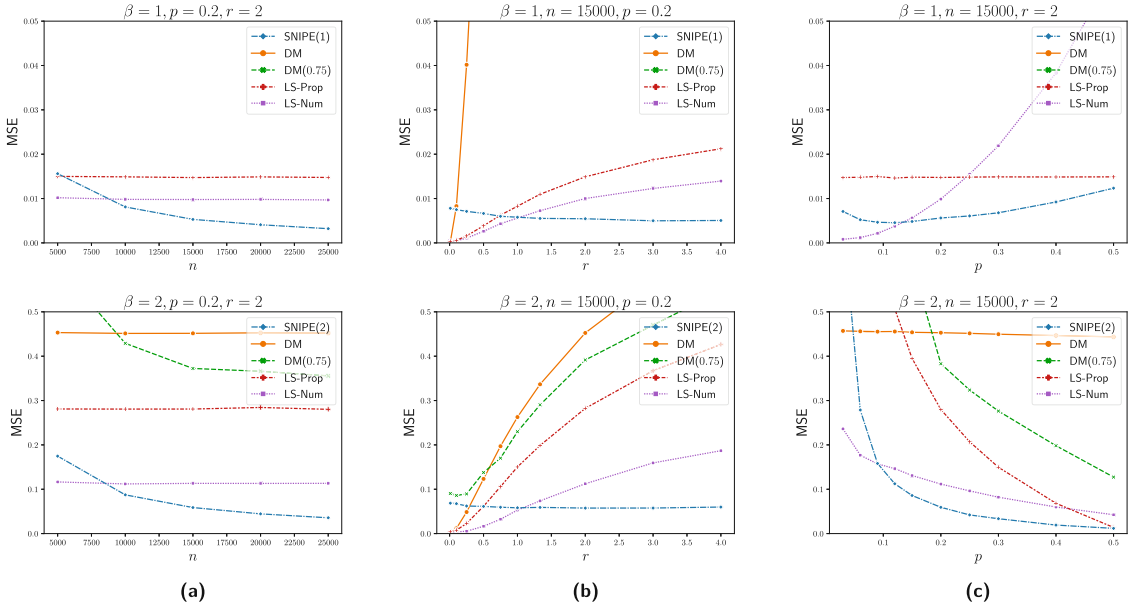
We also compare with least-squares regression models of degree β, which assume that the potential outcomes model is given by

$$Y_i(\mathbf{z}) = g(z_i, \bar{z}_i) = \left(\rho + \sum_{k=1}^{\beta} \gamma_k X_i^k\right) + z_i\left(\tilde{\rho} + \sum_{k=1}^{\beta-1} \tilde{\gamma}_k X_i^k\right), \tag{6.4}$$

for some covariate X_i. We consider two variations. In the first, we set X_i equal to the number of treated neighbors. In the second, we let X_i equal the proportion of treated neighbors. In both cases, we do not include i in its neighborhood. The two sets of coefficients $(\rho, \gamma_1, \ldots \gamma_\beta)$ and $(\tilde{\rho}, \tilde{\gamma}_1, \ldots \tilde{\gamma}_\beta)$ allow for the model to be different when i is treated vs not treated, and since we only allow up to degree β interactions, the second summation stops at $\beta - 1$. Overall, there are $2\beta + 1$ coefficients in the model. By using least-squares regression, we determine the set of coefficients minimizing the least-squares predictive error on the data set $\{z_i, X_i, Y_i(\mathbf{z})\}_{i\in[n]}$. These coefficients define an estimate for the function $\hat{g}$ in Equation 6.4. When $X_i = U_i$, the number of treated neighbors, the estimate is given by

$$\widehat{\mathrm{TTE}}_{\mathrm{LS-Num}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{g}(1, |\mathcal{N}_i| - 1) - \hat{g}(0,0)). \tag{6.5}$$

When we set $X_i = U_i/(|\mathcal{N}_i| - 1)$, the proportion of treated neighbors, we have



(a) (b) (c)

Figure 2: Plots visualizing the MSE of various TTE estimators under Bernoulli design on Erdős-Rényi networks for both linear and quadratic potential outcomes models. The height of each line on a plot depicts the mean squared error when the model is normalized so that the true TTE is effectively equal to 1. Alternatively, we can think of this as the variance of the normalized estimates. Our estimator under a β-order potential outcomes model is denoted SNIPE(β) in the figure. (a) Varying population size, (b) varying direct:indirect effects, and (c) varying treatment budget.

$$\widehat{\mathrm{TTE}}_{\mathrm{LS-Prop}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{g}(1,1) - \hat{g}(0,0)). \tag{6.6}$$

6.2 Results and discussion

For each population size n, we sample G networks from the Erdös–Rényi model described previously. For every configuration of parameters in the experiment, we sample N treatment assignment vectors $\mathbf{z}_1, \ldots, \mathbf{z}_N$ from a uniform Bernoulli distribution with treatment probability p to compute the TTE using each estimator. Each plot we include also shows the relative bias of the TTE estimates, averaged over the results from these GN samples and normalized by the magnitude, for each estimator. The width of the shading around each line in the plots shows the standard deviation across the GN estimates. For our experiments,[1] we chose $G = 10$ and $N = 500$.

Figures 1 and 2 visualize the effects of various network or estimator parameters on the performance of each of the four TTE estimators described in Section 6.1 and $\widehat{\mathrm{TTE}}_{\mathrm{SNIPE}(\beta)}$, all under Bernoulli RD. In particular, we consider the effects of the population size (n), the treatment budget (p), the ratio between the network and direct effects (r), and the degree of the potential outcomes model (β). We list specific values for the parameters above each plot. Figure 1 shows the bias and empirical standard deviation of each estimator, where the values are all normalized by the magnitude of the true TTE. Figure 2 plots the empirical MSE of each estimator, also normalized by the magnitude of the true TTE. The normalization can alternately be viewed as standardizing all models so that the ground truth TTE is 1.

The top row of plots in Figure 1 features results for a linear ($\beta = 1$) potential outcomes model while the bottom row shows results for a quadratic ($\beta = 2$) potential outcomes model. As expected, the SNIPE estimator,

1 The Python scripts for the experiments and the data used in our results are available at: https://github.com/mayscortez/low-order-unitRD.

Table 2: Table presenting the empirical variance of the SNIPE(β) estimator for the TTE under Bernoulli(p) design on Erdös–Rényi networks with $\beta = 1, 2$

	Experimental variance	Variance estimate	Variance bound
Results for SNIPE(1)			
n ($p = 0.2, r = 2$)			
1,000	17.63	16327.43	488300.11
2,500	8.19	3724.33	245279.16
5,000	3.34	2270.81	140688.10
7,500	2.33	1408.50	106981.33
10,000	1.95	1437.32	105354.62
p ($n = 5{,}000, r = 2$)			
0.1	2.93	5439.83	272939.86
0.2	3.71	2375.13	174599.20
0.3	4.33	1259.70	106516.48
0.4	6.10	1660.36	122405.92
0.5	8.06	2197.06	93585.34
r ($n = 5{,}000, p = 0.2$)			
0.5	1.06	1480.23	19410.59
1.0	1.69	1793.46	44527.96
1.5	2.50	3802.44	112805.72
2.0	3.91	2080.84	142140.47
Results for SNIPE(2)			
n ($p = 0.2, r = 2$)			
1,000	255.59	11046.03	470718.02
2,500	92.21	3873.10	231156.58
5,000	46.36	2962.08	147815.75
7,500	29.80	1495.31	95530.25
10,000	21.28	1463.03	81113.89
p ($n = 5{,}000, r = 2$)			
0.1	114.95	8635.44	147108.70
0.2	43.94	2142.06	133774.85
0.3	24.36	1373.10	129746.07
0.4	13.24	1115.97	128101.61
0.5	9.19	2435.29	133271.42
r ($n = 5{,}000, p = 0.2$)			
0.5	12.99	1627.89	126414.68
1.0	20.32	1731.70	129161.00
1.5	30.94	2054.55	134921.48
2.0	44.50	2494.29	144292.22

The variance bound is computed using the bound in Theorem 1 and the variance estimate is computed using the Aronow–Samii estimator described in section 5.3. The parameters n, p, and r refer to the population size, the treatment probability, and the ratio of direct to indirect effects, respectively.

shown in blue, has no relative bias and its variance decreases as n increases. With the exception of the modified difference-in-means estimator $\widehat{\text{TTE}}_{\text{DM}(0.75)}$ in green, the variances of the other estimators are lower than ours. However, the biases of the other estimators are larger than the standard deviation of our unbiased estimator overall. Moreover, as r increases, the networks effects are more significant than the direct effects and we see the biases of the other estimators grow larger. Note that the variance of our estimator remains relatively constant as r varies. When r is close to 0, there are essentially no network effects, SUTVA holds, and as expected, all the estimators are unbiased. Figure 2 shows that for many parameter combinations, the MSE of our estimator is lower than the other estimators; this is particularly the case for sufficiently large population sizes (large n) and sufficiently significant relative network effects (large r). In the top row of plots, corresponding to $\beta = 1$, the difference in means estimators perform poorly relative to the other estimators so that they are beyond the upper limit of the displayed plot. While the performance of the least squares estimators

and our estimator is comparable for $\beta = 1$, the MSE of our estimator is solely due to variance, which will decrease with large n, yet the MSE of the least squares estimators is largely due to its bias, which will not decrease with large n, highlighting that they are not consistent estimators for our heterogeneous model.

6.3 Variance estimator experiments

We compare the empirical variance of SNIPE with the variance bound from Theorem 1 and the variance estimator constructed from Aronow–Samii's method described in Section 5.3. As mentioned earlier, for each population size n, we sample G networks from the Erdös–Rényi model described previously. For every configuration of parameters in the experiment, we sample N treatment assignment vectors $\mathbf{z}_1, \ldots, \mathbf{z}_N$ from a uniform Bernoulli distribution with treatment probability p to compute the TTE using each estimator. For the variance experiments, we chose $G = 10$ and $N = 100$. Table 2 displays the effects of various network or estimator parameters on the experimental variance of SNIPE(β) as well as the variance estimate and the theoretical variance bound, all under Bernoulli RD, and all averaged over GN samples of treatment vectors. As in previous experiments, we consider the effects of the population size (n), the treatment budget (p), the ratio between the network and direct effects (r), and the degree of the potential outcomes model (β). We list fixed values for the parameters in parentheses. The main observation we wish to draw attention to is the differences in the orders of magnitude among the empirical variance, the variance estimate, and the variance bound. It is clear from these results that obtaining a tighter variance estimator would be a valuable direction for the future work.

7 Conclusions and future work

We propose an estimator for the TTE under neighborhood interference and Bernoulli design when the graph is known. Our approach considers a potential outcomes model that is polynomial in the treatment vector $\mathbf{z}$ with degree parameterized by β, which we assume to be much smaller than the maximum neighborhood size. This assumption is equivalent to constraining the order of interactions amongst treated neighbors to sets of size at most β. We derive theoretical bounds on the variance of our estimator under Bernoulli RD and show that we improve upon the variance of the Horvitz–Thompson estimator when β is significantly lower than the maximum degree of the graph. We provide minimax lower bounds on the mean squared error of our estimator when the graph is d-regular and the treatment probabilities are the same for each individual. Furthermore, under additional boundedness conditions, we prove a central limit theorem for our estimator, allowing for conservative, asymptotically valid confidence intervals using our proposed variance estimator. Through computational experiments, we illustrate that our estimator has lower MSE than the MSE of standard difference in means and least-squares estimators for the TTE. Our work uniquely complements the literature in that we consider how to incorporate and exploit structure in the potential outcomes model in a way that allows for a richer model class than the typical parametric model classes and does not reduce the effective treatment to a low-dimensional statistic.

Our work presents many interesting and likely fruitful directions for future work. We summarize a few of these in the following section.

7.1 Optimized experimental designs

In this work, we studied the relationship between the complexity of the potential outcomes model (parameterized by β) and the difficulty of estimation. In this analysis, we made no structural assumptions on the network and restricted focus to independent Bernoulli experimental design. It is easy to conceive that a more careful selection of the experimental design, motivated by structural information of the causal network, could lead to improved performance of the estimator. For example, in graphs that are well clustered, correlating the

treatments within each cluster could allow some individuals to have more of their neighborhood treated, giving a better estimate of the magnitude of their treatment effect. The design philosophy of our estimator, as motivated in Section 4.1 through the lens of experiment replication, is not particular to Bernoulli design. As such, an enticing direction of future study would be to explore this estimator for other experimental designs and better understand the interplay between performance gains due to the network structure and the model structure.

7.2 Implications to observational studies

While our stated theoretical results only hold for nonuniform Bernoulli designs, there are natural implications to the analysis of observational studies under appropriate unconfoundedness assumptions. In particular, if treatments across individuals are independent from each other conditioned on observed covariates, and if the conditional treatment probabilities could be estimated, then one could plausibly consider a plugin approach to modify our estimator for such observational data. Formalizing how to extend our results to observational settings would be a fruitful and interesting direction for the future work.

7.3 Dealing with model misspecification

Another interesting direction for future work centers around how to use our proposed class of estimators when the model parameter β is unknown. In settings such as online social networks, it is reasonable to posit a low-degree interactions assumption on the network interference, which corresponds to adopting a potential outcomes model parameterized by a ground-truth value β^{GT}. To estimate the TTE, a researcher will select a value β^{Exp} to use in defining their estimator. Without knowledge of the ground truth model, it is possible that $\beta^{\mathrm{GT}} \neq \beta^{\mathrm{Exp}}$. As this phenomenon of *model misspecification* is pervasive through causal inference and more general machine learning domains, it would be useful to quantify the relationship between the degree of β-misspecification and any additional accrual of bias or variance. Another related question is whether a statistical test can be developed to aid in the correct choice of β^{Exp} or to validate the low-degree polynomial structure of the model.

Acknowledgements: We gratefully acknowledge the financial support from the National Science Foundation grants CCF-1948256 and CNS-1955997 and the National Science Foundation Graduate Research Fellowship grant DGE-1650441. We also acknowledge support from the Air Force Research Laboratory grant AFOSR FA9550-23-1-0301. We also thank Professor Nathan Kallus for his insightful feedback.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: The authors state no conflict of interest.

Data availability statement: The Python scripts for the experiments and the data used in our results are available at: https://github.com/mayscortez/low-order-unitRD.

References

[1] Ugander J, Yin H. Randomized graph cluster randomization. 2020. https://arxiv.org/abs/2009.02297.

[2] Rubin DB. Randomization analysis of experimental data: the fisher randomization test comment. J Amer Stat Assoc. 1980;75(371):591–3. http://www.jstor.org/stable/2287653.

[3] Aronow PM, Samii C. Estimating average causal effects under general interference, with application to a social network experiment. Ann Appl Stat. 2017;11(4):1912–47.
[4] Manski CF. Identification of treatment response with social interactions. Econom J. 2013;16(1):S1–23.
[5] Basse GW, Airoldi EM. Limitations of design-based causal inference and A/B testing under arbitrary and network interference. Sociol Methodol. 2018;48(1):136–51.
[6] Toulis P, Kao E. Estimation of causal peer influence effects. In: International Conference on Machine Learning; 2013. p. 1489–97.
[7] Gui H, Xu Y, Bhasin A, Han J. Network a/b testing: From sampling to estimation. In: Proceedings of the 24th International Conference on International Conferences Steering Committee; 2015. p. 399–409.
[8] Basse GW, Airoldi EM. Model-assisted design of experiments in the presence of network-correlated outcomes. Biometrika. 2018;105(4):849–58.
[9] Cai J, De Janvry A, Sadoulet E. Social networks and the decision to insure. Amer Econ J Appl Econ. 2015;7(2):81–108.
[10] Parker BM, Gilmour SG, Schormans J. Optimal design of experiments on connected units with application to social networks. J R Stat Soc Ser C (Appl Stat.) 2017;66(3):455–80. https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssc.12170.
[11] Chin A. Regression adjustments for estimating the global treatment effect in experiments with interference. J Causal Inference. 2019;7(2):20180026.
[12] Sobel ME. What do randomized studies of housing mobility demonstrate? J Amer Stat Assoc. 2006;101(476):1398–407. doi: 10.1198/016214506000000636.
[13] Rosenbaum PR. Interference between units in randomized experiments. J Amer Stat Assoc. 2007;102(477):191–200. doi: 10.1198/016214506000001112.
[14] Hudgens MG, Halloran ME. Toward causal inference with interference. J Amer Stat Assoc. 2008;103:832–42. https://EconPapers.repec.org/RePEc:bes:jnlasa:v:103:y:2008:m:june:p:832-842.
[15] Tchetgen EJT, VanderWeele TJ. On causal inference in the presence of interference. Stat Meth Med Res. 2012;21(1):55–75. PMID: 21068053. doi: 10.1177/0962280210386779.
[16] Eckles D, Karrer B, Ugander J. Design and analysis of experiments in networks: reducing bias from interference. J Causal Inference. 2017;5(1):20150021.
[17] Ugander J, Karrer B, Backstrom L, Kleinberg J. Graph cluster randomization: Network exposure to multiple universes. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM; 2013. p. 329–37.
[18] Sussman DL, Airoldi EM. Elements of estimation theory for causal effects in the presence of network interference. 2017. http://arXiv.org/abs/arXiv:170203578.
[19] Bargagli-Stoffi FJ, Tortù C, Forastiere L. Heterogeneous treatment and spillover effects under clustered network interference. 2020. http://arXiv.org/abs/arXiv:200800707.
[20] Bhattacharya R, Malinsky D, Shpitser I. Causal inference under interference and network uncertainty. In: Adams RP, Gogate V, editors. Proceedings of The 35th Uncertainty in Artificial Intelligence Conference. vol. 115 of Proceedings of Machine Learning Research. PMLR; 2020. p. 1028–38. https://proceedings.mlr.press/v115/bhattacharya20a.html.
[21] Liu L, Hudgens MG. Large sample randomization inference of causal effects in the presence of interference. J Amer Stat Assoc. 2014;109(505):288–301. PMID: 24659836. doi: 10.1080/01621459.2013.844698.
[22] Li S, Wager S. Random graph asymptotics for treatment effect estimation under network interference. 2020. https://arxiv.org/abs/2007.13302.
[23] Viviano D. Experimental design under network interference. 2020. https://arxiv.org/abs/2003.08421.
[24] Leung MP. Rate-optimal cluster-randomized designs for spatial interference. 2021. https://arxiv.org/abs/2111.04219.
[25] Athey S, Eckles D, Imbens GW. Exact p-values for network interference. J Amer Stat Assoc. 2018;113(521):230–40.
[26] VanderWeele TJ, TchetgenTchetgen EJ, Halloran ME. Interference and sensitivity analysis. Statist Sci. 2014 Nov;29(4):687–706. doi: 10.1214/14-STS479.
[27] Auerbach E, Tabord-Meehan M. The local approach to causal inference under network interference. 2021. http://arXiv.org/abs/arXiv:210503810.
[28] Taljaard M, Weijer C, Grimshaw JM, BelleBrown J, Binik A, Boruch R, et al. Ethical and policy issues in cluster randomized trials: rationale and design of a mixed methods research study. Trials. 2009;10(1):1–10.
[29] Edwards SJ, Braunholtz DA, Lilford RJ, Stevens AJ. Ethical issues in the design and conduct of cluster randomised controlled trials. Bmj. 1999;318(7195):1407–9.
[30] Hutton JL. Are distinctive ethical principles required for cluster randomized controlled trials? Stat Med. 2001;20(3):473–88.
[31] Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. Amer J Public Health. 2004;94(3):416–22. PMID: 14998805. doi: 10.2105/AJPH.94.3.416.
[32] Johari R, Li H, Liskovich I, Weintraub GY. Experimental design in two-sided platforms: an analysis of bias. Manag Sci. 2022;68:7069–89.
[33] Li H, Zhao G, Johari R, Weintraub GY. Interference, bias, and variance in two-sided marketplace experimentation: guidance for platforms. In: Proceedings of the ACM Web Conference 2022; 2022. p. 182–92.
[34] Spang B, Hannan V, Kunamalla S, Huang TY, McKeown N, Johari R. Unbiased experiments in congested networks. In: Proceedings of the 21st ACM Internet Measurement Conference; 2021. p. 80–95.
[35] Bright I, Delarue A, Lobel I. Reducing marketplace interference bias via shadow prices. 2022. http://arXiv.org/abs/arXiv:220502274.

[36] Perez-Heydrich C, Hudgens MG, Halloran ME, Clemens JD, Ali M, Emch ME. Assessing effects of cholera vaccination in the presence of interference. Biometrics. 2014;70(3):731–41.
[37] Liu L, Hudgens MG, Becker-Dreps S. On inverse probability-weighted estimators in the presence of interference. Biometrika. 2016;103(4):829–42.
[38] DiTraglia FJ, Garcia-Jimeno C, O'Keeffe-O'Donovan R, Sanchez-Becerra A. Identifying causal effects in experiments with spillovers and non-compliance. 2020. https://arxiv.org/abs/2011.07051.
[39] Vazquez-Bare G. Identification and estimation of spillover effects in randomized experiments. J Econom. 2022;105237.
[40] Verbitsky-Savitz N, Raudenbush SW. Causal inference under interference in spatial settings: a case study evaluating community policing program in Chicago. Epidemiol Meth. 2012;1(1):107–30.
[41] Ogburn EL, Sofrygin O, Diaz I, Van der Laan MJ. Causal inference for social network data. 2017. http://arXiv.org/abs/arXiv:170508527.
[42] Forastiere L, Airoldi EM, Mealli F. Identification and estimation of treatment and interference effects in observational studies on networks. J Amer Stat Assoc. 2021;116(534):901–18.
[43] Cortez M, Eichhorn M, Yu CL. Staggered rollout designs enable causal inference under interference without network knowledge. In: Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K, Oh A, editors. Advances in Neural Information Processing Systems. Curran Associates, Inc.; Vol. 35. 2022.
[44] Yu CL, Airoldi E, Borgs C, Chayes J. Estimating the total treatment effect in randomized experiments with unknown network structure. Proceedings of the National Academy of Sciences. 2022;119(44):e2208975119.
[45] Swaminathan A, Krishnamurthy A, Agarwal A, Dudik M, Langford J, Jose D, et al. Off-policy evaluation for slate recommendation. In: Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. Advances in Neural Information Processing Systems. Curran Associates, Inc.; Vol. 30. 2017.
[46] Harshaw C, Sävje F, Wang Y. A design-based Riesz representation framework for randomized experiments. 2022. http://arXiv.org/abs/arXiv:221008698.
[47] LeCam L. Convergence of estimates under dimensionality restrictions. Ann Stat. 1973;1:38–53.
[48] Tsybakov AB. Introduction to nonparametric estimation. New York: Springer; 2009.
[49] Chin A. Central limit theorems via Stein's method for randomized experiments under interference. 2018. http://arXiv.org/abs/arXiv:180403105.
[50] Leung MP. Treatment and spillover effects under network interference. Rev Econ Stat. 2020 May;102(2):368–80.
[51] Harshaw C, Sävje F, Eisenstat D, Mirrokni V, Pouget-Abadie J. Design and analysis of bipartite experiments under a linear exposure-response model. Electr J Stat. 2023;17(1):464–518.
[52] Ogburn EL, Sofrygin O, Diaz I, Van der Laan MJ. Causal inference for social network data. J Amer Stat Assoc. 2022:1–15.
[53] Ross N. Fundamentals of Steinas method. Probabil Surveys. 2011;8:210–93.
[54] Leung MP. Rate-optimal cluster-randomized designs for spatial interference. Ann Stat. 2022;50(5):3064–87.
[55] Aronow PM, Samii C. Conservative variance estimation for sampling designs with zero pairwise inclusion probabilities. Survey Methodol. 2013;39(1):231–41.

Appendix

A The explicit TTE estimator for general β

Here, we derive an explicit formula for the TTE estimator under non-uniform Bernoulli design for general β. Recall (Equation (4.3)) that the estimator has the following form:

$$\widehat{\mathrm{TTE}} = \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z})\langle \mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]^{-1}(\mathbf{1}_{|\mathcal{S}_i^\beta|} - \mathbf{e}_1), \tilde{\mathbf{z}}_i\rangle,$$

where each $\mathcal{S}_i^\beta$ collects all subsets of $\mathcal{N}_i$ with cardinality at most β and each $\tilde{\mathbf{z}}_i$ is the vector of length $|S_i|$ with entries $(\tilde{\mathbf{z}}_i)_{\mathcal{S}} = \prod_{j\in\mathcal{S}} z_j$, indicating whether the entire subset $\mathcal{S}$ has been assigned to treatment. Here, we index vectors and matrix entries by their corresponding sets (rather than numerical indices) for notational convenience. To begin, we focus our attention on the first argument vectors of these inner products. Entries of the matrix $\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]$ have the form

$$(\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top])_{\mathcal{S},\mathcal{T}} = \mathbb{E}\left[\prod_{j\in\mathcal{S}} z_j \prod_{j'\in\mathcal{T}} z_{j'}\right] = \mathbb{E}\left[\prod_{j\in\mathcal{S}\cup\mathcal{T}} z_j\right] = \prod_{j\in\mathcal{S}\cup\mathcal{T}} p_j.$$

Here, the second equality uses the fact that each $z_j \in \{0,1\}$, and the third equality uses the independence of the treatment assignments. The following lemma establishes the invertibility of this matrix by giving an explicit expression for its inverse.

Lemma 1. *The matrix $\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]$ is invertible, with entries of its inverse A_i given by the formula*

$$(A_i)_{\mathcal{S},\mathcal{T}} = \prod_{j\in\mathcal{S}}\frac{-1}{p_j}\prod_{k\in\mathcal{T}}\frac{-1}{p_k}\sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta\\(\mathcal{S}\cup\mathcal{T})\subseteq\mathcal{U}}}\prod_{\ell\in\mathcal{U}}\frac{p_\ell}{1-p_\ell}.$$

Proof. We will argue that $\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]A_i = I_{|\mathcal{S}_i^\beta|}$ entrywise. Note that both $\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]$ and A_i are symmetric matrices, so their product will be as well. First, we consider the diagonal entries of this matrix. Given any $\mathcal{S}\in\mathcal{S}_i^\beta$,

$$\begin{aligned}
(\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]A_i)_{\mathcal{S},\mathcal{S}} &= \sum_{\mathcal{T}\in\mathcal{S}_i^\beta}\prod_{j\in\mathcal{S}\cup\mathcal{T}} p_j\,(A_i)_{\mathcal{T},\mathcal{S}}\\
&= \sum_{\mathcal{T}\in\mathcal{S}_i^\beta}(-1)^{|\mathcal{S}\cup\mathcal{T}|}\prod_{k\in(\mathcal{S}\cap\mathcal{T})}\frac{-1}{p_k}\sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta\\(\mathcal{S}\cup\mathcal{T})\subseteq\mathcal{U}}}\prod_{\ell\in\mathcal{U}}\frac{p_\ell}{1-p_\ell}\\
&= \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta\\\mathcal{S}\subseteq\mathcal{U}}}\prod_{\ell\in\mathcal{U}}\frac{p_\ell}{1-p_\ell}\sum_{\mathcal{T}\subseteq\mathcal{U}}(-1)^{|\mathcal{S}\cup\mathcal{T}|}\prod_{k\in(\mathcal{S}\cap\mathcal{T})}\frac{-1}{p_k} && \text{(reverse sum)}\\
&= (-1)^{|\mathcal{S}|}\sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta\\\mathcal{S}\subseteq\mathcal{U}}}\prod_{\ell\in\mathcal{U}}\frac{p_\ell}{1-p_\ell}\sum_{\mathcal{V}\subseteq\mathcal{S}}\prod_{k\in\mathcal{V}}\frac{-1}{p_k}\underbrace{\sum_{\mathcal{W}\subseteq(\mathcal{U}\setminus\mathcal{S})}(-1)^{|\mathcal{W}|}}_{\mathbb{1}(\mathcal{U}\subseteq\mathcal{S})} && (\mathcal{V}=\mathcal{T}\cap\mathcal{S},\ \mathcal{W}=\mathcal{T}\setminus\mathcal{S})\\
&= (-1)^{|\mathcal{S}|}\prod_{\ell\in\mathcal{S}}\frac{p_\ell}{1-p_\ell}\sum_{\mathcal{V}\subseteq\mathcal{S}}\prod_{k\in\mathcal{V}}\frac{-1}{p_k} && \text{(only non-zero term is } \mathcal{U}=\mathcal{S})\\
&= (-1)^{|\mathcal{S}|}\prod_{\ell\in\mathcal{S}}\frac{p_\ell}{1-p_\ell}\prod_{k\in\mathcal{S}}\left(1-\frac{1}{p_k}\right) && \text{(distributivity)}\\
&= 1.
\end{aligned}$$

Next, we consider the off-diagonal entries. By the symmetry of $\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]A_i$, it suffices to consider entries $(\mathcal{S}',\mathcal{S})$ for which $\mathcal{S}'\setminus\mathcal{S}\neq\emptyset$. We have,

$$
\begin{aligned}
(M_i A_i)_{\mathcal{S}',\mathcal{S}} &= \sum_{\mathcal{T}\in\mathcal{S}_i^\beta} \prod_{j'\in\mathcal{S}'\cup\mathcal{T}} p_{j'}\,(A_i)_{\mathcal{T},\mathcal{S}} \\
&= \sum_{\mathcal{T}\in\mathcal{S}_i^\beta} \prod_{j'\in\mathcal{S}'\cup\mathcal{T}} p_{j'} \prod_{j\in\mathcal{S}} \frac{-1}{p_j} \prod_{k\in\mathcal{T}} \frac{-1}{p_k} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta \\ (\mathcal{S}\cup\mathcal{T})\subseteq\mathcal{U}}} \prod_{\ell\in\mathcal{U}} \frac{p_\ell}{1-p_\ell} \\
&= \prod_{j\in\mathcal{S}} \frac{-1}{p_j} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta \\ \mathcal{S}\subseteq\mathcal{U}}} \prod_{\ell\in\mathcal{U}} \frac{p_\ell}{1-p_\ell} \sum_{\mathcal{T}\subseteq\mathcal{U}} \prod_{j'\in\mathcal{S}'\cup\mathcal{T}} p_{j'} \prod_{k\in\mathcal{T}} \frac{-1}{p_k} && \text{(reverse sum)} \\
&= \prod_{j\in\mathcal{S}} \frac{-1}{p_j} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta \\ \mathcal{S}\subseteq\mathcal{U}}} \prod_{\ell\in\mathcal{U}} \frac{p_\ell}{1-p_\ell} \prod_{j'\in\mathcal{S}'} p_{j'} \sum_{\mathcal{V}\subseteq\mathcal{S}'} \prod_{k\in\mathcal{V}} \frac{-1}{p_k} \underbrace{\sum_{\mathcal{W}\subseteq\mathcal{U}\setminus\mathcal{S}'} (-1)^{|\mathcal{W}|}}_{\mathbb{1}(\mathcal{U}\subseteq\mathcal{S}')} \quad (\mathcal{V}=\mathcal{T}\cap\mathcal{S}',\ \mathcal{W}=\mathcal{T}\setminus\mathcal{S}') \\
&= 0.
\end{aligned}
$$

Here, the last line follows because any non-zero term in the outer sum must correspond to some $\mathcal{U}$ such that $\mathcal{S}\subseteq\mathcal{U}\subseteq\mathcal{S}' \Rightarrow \mathcal{S}\subseteq\mathcal{S}'$. Our earlier assumption that $\mathcal{S}'\setminus\mathcal{S}\neq\varnothing$ ensures there is no such $\mathcal{U}$. □

By using this lemma, we consider the entries in the first argument vector of each inner product. We have,

$$
\begin{aligned}
(A_i(\mathbf{1}_{|\mathcal{S}_i^\beta|}-\mathbf{e}_1))_{\mathcal{S}} &= \sum_{\mathcal{T}\in\mathcal{S}_i^\beta} (A_i)_{\mathcal{S},\mathcal{T}} - (A_i)_{\mathcal{S},\varnothing} \\
&= \prod_{j\in\mathcal{S}} \frac{-1}{p_j} \left[\sum_{\mathcal{T}\in\mathcal{S}_i^\beta} \prod_{k\in\mathcal{T}} \frac{-1}{p_k} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta \\ (\mathcal{S}\cup\mathcal{T})\subseteq\mathcal{U}}} \prod_{\ell\in\mathcal{U}} \frac{p_\ell}{1-p_\ell} - \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta \\ \mathcal{S}\subseteq\mathcal{U}}} \prod_{\ell\in\mathcal{U}} \frac{p_\ell}{1-p_\ell} \right] \\
&= \prod_{j\in\mathcal{S}} \frac{-1}{p_j} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta \\ \mathcal{S}\subseteq\mathcal{U}}} \prod_{\ell\in\mathcal{U}} \frac{p_\ell}{1-p_\ell} \left[\sum_{\mathcal{T}\subseteq\mathcal{U}} \prod_{k\in\mathcal{T}} \frac{-1}{p_k} - 1 \right] && \text{(reverse sum and factor)} \\
&= \prod_{j\in\mathcal{S}} \frac{-1}{p_j} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta \\ \mathcal{S}\subseteq\mathcal{U}}} \prod_{\ell\in\mathcal{U}} \frac{p_\ell}{1-p_\ell} \left[\prod_{k\in\mathcal{U}} \frac{p_k-1}{p_k} - 1 \right] && \text{(distributivity)} \\
&= \prod_{j\in\mathcal{S}} \frac{-1}{p_j} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta \\ \mathcal{S}\subseteq\mathcal{U}}} g(\mathcal{U}) \prod_{\ell\in\mathcal{U}} \frac{-1}{1-p_\ell},
\end{aligned}
$$

where we define $g(\mathcal{S}) \coloneqq \prod_{s\in\mathcal{S}}(1-p_s) - \prod_{s\in\mathcal{S}}(-p_s)$. By substituting back into the inner product, we calculate

$$
\begin{aligned}
\langle A_i(\mathbf{1}_{|\mathcal{S}_i^\beta|}-\mathbf{e}_1), \tilde{\mathbf{z}}_i\rangle &= \sum_{\mathcal{S}\in\mathcal{S}_i^\beta} \prod_{j\in\mathcal{S}} \frac{-z_j}{p_j} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta \\ \mathcal{S}\subseteq\mathcal{U}}} g(\mathcal{U}) \prod_{\ell\in\mathcal{U}} \frac{-1}{1-p_\ell} \\
&= \sum_{\mathcal{U}\in\mathcal{S}_i^\beta} g(\mathcal{U}) \prod_{\ell\in\mathcal{U}} \frac{-1}{1-p_\ell} \sum_{\mathcal{S}\subseteq\mathcal{U}} \prod_{j\in\mathcal{S}} \frac{-z_j}{p_j} \\
&= \sum_{\mathcal{U}\in\mathcal{S}_i^\beta} g(\mathcal{U}) \prod_{\ell\in\mathcal{U}} \frac{-1}{1-p_\ell} \prod_{j\in\mathcal{U}} \left(1-\frac{z_j}{p_j}\right) \\
&= \sum_{\mathcal{U}\in\mathcal{S}_i^\beta} g(\mathcal{U}) \prod_{j\in\mathcal{U}} \frac{z_j-p_j}{p_j(1-p_j)}.
\end{aligned}
$$

Finally, we replace $\mathcal{U}$ with $\mathcal{S}$ to conform to earlier notation and obtain the explicit form for our estimator

$$
\widehat{TTE} = \frac{1}{n}\sum_{i=1}^{n} Y_i \sum_{\mathcal{S}\in\mathcal{S}_i^\beta} g(\mathcal{S}) \prod_{j\in\mathcal{S}} \frac{z_j-p_j}{p_j(1-p_j)}.
$$

B Proof of Theorem 1

B.1 Unbiasedness

The key insight that we use in our unbiasedness calculations comes from the following lemma.

Lemma 2. *If* $\{z_j\}_{j\in[n]}$ *are mutually independent with* $z_j \sim \text{Bernoulli}(p_j)$*, then for any* $\mathcal{S}, \mathcal{S}' \subseteq [n]$*,*

$$\mathbb{E}\left[\prod_{j\in\mathcal{S}} \frac{z_j - p_j}{p_j(1-p_j)} \prod_{j'\in\mathcal{S}'} z_{j'}\right] = \mathbb{I}(\mathcal{S}\subseteq\mathcal{S}')\cdot \prod_{j'\in\mathcal{S}'\setminus\mathcal{S}} p_{j'}.$$

Proof. By the mutual independence of the $\{z_j\}$, we can rewrite this expectations as a product, separating the variables into three groups.

$$\mathbb{E}\left[\prod_{j\in\mathcal{S}} \frac{z_j - p_j}{p_j(1-p_j)} \prod_{j'\in\mathcal{S}'} z_{j'}\right] = \prod_{j\in\mathcal{S}\setminus\mathcal{S}'} \mathbb{E}\left[\frac{z_j - p_j}{p_j(1-p_j)}\right] \prod_{j'\in\mathcal{S}'\setminus\mathcal{S}} \mathbb{E}[z_{j'}] \prod_{j''\in\mathcal{S}\cap\mathcal{S}'} \mathbb{E}\left[\frac{z_{j''}(z_{j''} - p_{j''})}{p_{j''}(1-p_{j''})}\right].$$

Note that the expectations in the first product each simplify to 0, so this expectation is non-zero only when $\mathcal{S}\subseteq\mathcal{S}'$. The expectations in the second product simplify to $p_{j'}$, and those in the third product each simplify to 1. These observations imply the lemma. □

The critical feature of this lemma is that this indicator function simplifies sums over arbitrary sets to sums over subsets $\mathcal{S}\subseteq\mathcal{S}'$. This additional structure permits simplifications using the distributive property

$$\sum_{\mathcal{S}\subseteq\mathcal{S}'} \prod_{j\in\mathcal{S}} a_j = \prod_{j\in\mathcal{S}'} (1 + a_j).$$

We leverage this fact in the following calculation. Given any $\mathcal{S}' \in \mathcal{S}_i^\beta$, we may simplify

$$\begin{aligned}
\mathbb{E}\left[\sum_{\mathcal{S}\in\mathcal{S}_i^\beta} g(\mathcal{S}) \prod_{j\in\mathcal{S}} \frac{z_j - p_j}{p_j(1-p_j)} \prod_{j'\in\mathcal{S}'} z_{j'}\right] &= \sum_{\mathcal{S}\in\mathcal{S}_i^\beta} g(\mathcal{S})\cdot \mathbb{E}\left[\prod_{j\in\mathcal{S}} \frac{z_j - p_j}{p_j(1-p_j)} \prod_{j'\in\mathcal{S}'} z_{j'}\right] && \text{(linearity)}\\
&= \sum_{\mathcal{S}\subseteq\mathcal{S}'} g(\mathcal{S}) \prod_{j'\in\mathcal{S}'\setminus\mathcal{S}} p_{j'} && \text{(Lemma 2)}\\
&= \prod_{j'\in\mathcal{S}'} p_{j'} \sum_{\mathcal{S}\subseteq\mathcal{S}'} g(\mathcal{S}) \prod_{j\in\mathcal{S}} \frac{1}{p_j}\\
&= \prod_{j'\in\mathcal{S}'} p_{j'} \left(\sum_{\mathcal{S}\subseteq\mathcal{S}'} \prod_{j\in\mathcal{S}} \frac{1-p_j}{p_j} - \sum_{\mathcal{S}\subseteq\mathcal{S}'} \prod_{j\in\mathcal{S}} (-1)\right) && \text{(definition of } g(\mathcal{S}))\\
&= \prod_{j'\in\mathcal{S}'} p_{j'} \left(\prod_{j\in\mathcal{S}'} \left(1 + \frac{1-p_j}{p_j}\right) - \mathbb{I}(\mathcal{S}' = \varnothing)\right) && \text{(distributivity)}\\
&= 1 - \mathbb{I}(\mathcal{S}' = \varnothing)\\
&= \mathbb{I}(\mathcal{S}' \neq \varnothing).
\end{aligned}$$

By applying the linearity of expectation and the previous result, we calculate

$$\begin{aligned}
\mathbb{E}[\widehat{\text{TTE}}] &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[Y_i \sum_{\mathcal{S}\in\mathcal{S}_i^\beta} g(\mathcal{S}) \prod_{j\in\mathcal{S}} \frac{z_j - p_j}{p_j(1-p_j)}\right]\\
&= \frac{1}{n}\sum_{i=1}^n \sum_{\mathcal{S}'\in\mathcal{S}_i^\beta} c_{i,\mathcal{S}'}\, \mathbb{E}\left[\sum_{\mathcal{S}\in\mathcal{S}_i^\beta} g(\mathcal{S}) \prod_{j\in\mathcal{S}} \frac{z_j - p_j}{p_j(1-p_j)} \prod_{j'\in\mathcal{S}'} z_{j'}\right]\\
&= \frac{1}{n}\sum_{i=1}^n \sum_{\substack{\mathcal{S}'\in\mathcal{S}_i^\beta\\ \mathcal{S}'\neq\varnothing}} c_{i,\mathcal{S}'}\\
&= \text{TTE}.
\end{aligned}$$

B.2 Variance bound

To bound the variance of this estimator, we make use of the following lemma to bound the magnitude of each $g(\mathcal{S})$ coefficient.

Lemma 3. *For any* $\mathcal{S} \subseteq [n]$, $|g(\mathcal{S})| \leq 1$.

Proof. First, note that $|g(\varnothing)| = 0 \leq 1$. Now, for any non-empty set $\mathcal{S}$, let $i \in \mathcal{S}$. Then,

$$\begin{aligned}
|g(\mathcal{S})| &= \left| \prod_{s\in\mathcal{S}} (1-p_s) - \prod_{s\in\mathcal{S}} (-p_s) \right| \\
&= \left| (1-p_i) \prod_{s\in\mathcal{S}\setminus\{i\}} (1-p_s) + p_i \prod_{s\in\mathcal{S}\setminus\{i\}} (-p_s) \right| \\
&\leq (1-p_i) \prod_{s\in\mathcal{S}\setminus\{i\}} (1-p_s) + p_i \prod_{s\in\mathcal{S}\setminus\{i\}} p_s \qquad \text{(triangle inequality)} \\
&\leq 1 - p_i + p_i \\
&= 1.
\end{aligned}$$

This next lemma is used to bound the covariance terms that appear in our final calculation.

Lemma 4. *Suppose that* $\{z_j\}_{j\in[n]}$ *are mutually independent, with* $z_j \sim Bernoulli(p_j)$. *Then, for any* $\mathcal{S}, \mathcal{S}', \mathcal{T}, \mathcal{T}' \subseteq [n]$,

$$0 \leq Cov\left[\prod_{j\in\mathcal{S}} \frac{z_j - p_j}{p_j(1-p_j)} \prod_{j'\in\mathcal{S}'} z_{j'}, \prod_{k\in\mathcal{T}} \frac{z_k - p_k}{p_k(1-p_k)} \prod_{k'\in\mathcal{T}'} z_{k'} \right] \leq \mathbb{I}(\mathcal{S}\triangle\mathcal{T} \subseteq \mathcal{S}' \cup \mathcal{T}') \cdot \left(\frac{1}{p(1-p)} \right)^{|\mathcal{S}\cap\mathcal{T}|},$$

where $\mathcal{S}\triangle\mathcal{T} = (\mathcal{S}\cup\mathcal{T})\setminus(\mathcal{S}\cap\mathcal{T})$ *indicates the symmetric difference of* $\mathcal{S}$ *and* $\mathcal{T}$.

Proof. We reason separately about the two terms in the covariance expansion. By Lemma 2,

$$\mathbb{E}\left[\prod_{j\in\mathcal{S}} \frac{z_j - p_j}{p_j(1-p_j)} \prod_{j'\in\mathcal{S}'} z_{j'} \right] \mathbb{E}\left[\prod_{k\in\mathcal{T}} \frac{z_k - p_k}{p_k(1-p_k)} \prod_{k'\in\mathcal{T}'} z_{k'} \right] = \mathbb{I}\begin{pmatrix} \mathcal{S}\subseteq\mathcal{S}', \\ \mathcal{T}\subseteq\mathcal{T}' \end{pmatrix} \prod_{j'\in\mathcal{S}'\setminus\mathcal{S}} p_{j'} \prod_{k'\in\mathcal{T}'\setminus\mathcal{T}} p_{k'}. \tag{A1}$$

Next, we reason about the expectation of the product term. Since the z_j are Bernoulli random variables, we can combine the products over $\mathcal{S}'$ and $\mathcal{T}'$, giving

$$\mathbb{E}\left[\prod_{j\in\mathcal{S}} \frac{z_j - p_j}{p_j(1-p_j)} \prod_{k\in\mathcal{T}} \frac{z_k - p_k}{p_k(1-p_k)} \prod_{j'\in\mathcal{S}'\cup\mathcal{T}'} z_{j'} \right]. \tag{A2}$$

We partition the elements of $\mathcal{S}\cup\mathcal{S}'\cup\mathcal{T}\cup\mathcal{T}'$ based on which of the products they are present in:

(1) $j \in \mathcal{S}\cap\mathcal{T}\cap(\mathcal{S}'\cup\mathcal{T}')$: j contributes a factor of $\mathbb{E}\left[\frac{z_j^3 - 2z_j^2 p_j + z_j p_j^2}{p_j^2(1-p_j)^2}\right] = \frac{1}{p_j}$.

(2) $j \in \mathcal{S}\cap\mathcal{T}\setminus(\mathcal{S}'\cup\mathcal{T}')$: j contributes a factor of $\mathbb{E}\left[\frac{z_j^2 - 2z_j p_j + p_j^2}{p_j^2(1-p_j)^2}\right] = \frac{1}{p_j(1-p_j)}$.

(3) $j \in \mathcal{S}\cap(\mathcal{S}'\cup\mathcal{T}')\setminus\mathcal{T}$: j contributes a factor of $\mathbb{E}\left[\frac{z_j^2 - z_j p_j}{p_j - p_j^2}\right] = 1$.

(4) $j \in \mathcal{T}\cap(\mathcal{S}'\cup\mathcal{T}')\setminus\mathcal{S}$: j contributes a factor of $\mathbb{E}\left[\frac{z_j^2 - z_j p_j}{p_j - p_j^2}\right] = 1$.

(5) $j \in \mathcal{S}\setminus\mathcal{T}\setminus(\mathcal{S}'\cup\mathcal{T}')$: j contributes a factor of $\mathbb{E}\left[\frac{z_j - p_j}{p_j - p_j^2}\right] = 0$.

(6) $j \in \mathcal{T}\setminus\mathcal{S}\setminus(\mathcal{S}'\cup\mathcal{T}')$: j contributes a factor of $\mathbb{E}\left[\frac{z_j - p_j}{p_j - p_j^2}\right] = 0$.

(7) $j \in (\mathcal{S}'\cup\mathcal{T}')\setminus\mathcal{S}\setminus\mathcal{T}$: j contributes a factor of $\mathbb{E}[z_j] = p_j$.

Cases 5 and 6 ensure that (A2) is non-zero only when $\mathcal{S} \subseteq (\mathcal{T} \cup \mathcal{S}' \cup \mathcal{T}')$ and $\mathcal{T} \subseteq (\mathcal{S} \cup \mathcal{S}' \cup \mathcal{T}')$, or equivalently when $\mathcal{S} \triangle \mathcal{T} \subseteq \mathcal{S}' \cup \mathcal{T}'$. This condition is necessary for (A1) to be non-zero. In addition, note that each j from case 7 contributing a factor of p_j to (A2) also contributes at least one factor of p_j to (A1). The remaining j from other cases contribute a factor of at least 1. Notably, both (A2) and (A1) are non-negative, with (A2) dominating (A1), so that the covariance is always bounded below by zero, and upper bounded by (A2). In (A2), we can upper bound the contribution of each $j \in \mathcal{S} \cap \mathcal{T}$ by $(p(1-p))^{-1}$; note that our definition of p ensures that $p_j(1-p_j) \geq p(1-p)$ for each j. We upper bound the contribution of each other j by 1, which establishes the stated bound on the covariance. □

We are ready to bound the variance. If $\mathcal{N}_i \cap \mathcal{N}_{i'} = \emptyset$, then $Y_i(\mathbf{z})w_i(\mathbf{z})$ and $Y_{i'}(\mathbf{z})w_{i'}(\mathbf{z})$ are functions of disjoint sets of independent variables. Thus, $\mathrm{Cov}[Y_i(\mathbf{z})w_i(\mathbf{z}), Y_{i'}(\mathbf{z})w_{i'}(\mathbf{z})] = 0$. We let $\mathcal{M}_i$ denote the set of individuals i' such that $\mathcal{N}_i \cap \mathcal{N}_{i'} \neq \emptyset$, i.e., all individuals i' that share an in-neighbor with individual i. Note that $|\mathcal{M}_i| \leq d_{\text{in}} d_{\text{out}}$. By applying the bilinearity of covariance and the triangle inequality, we have

$$\mathrm{Var}[\widehat{\mathrm{TTE}}] \leq \frac{1}{n^2} \sum_{i=1}^{n} \sum_{i' \in \mathcal{M}_i} \sum_{\mathcal{S}' \in \mathcal{S}_i^\beta} |c_{i,\mathcal{S}'}| \sum_{\mathcal{T}' \in \mathcal{S}_{i'}^\beta} |c_{i',\mathcal{T}'}| \sum_{\mathcal{S} \in \mathcal{S}_i^\beta} |g(\mathcal{S})| \sum_{\mathcal{T} \in \mathcal{S}_{i'}^\beta} |g(\mathcal{T})| \times \left| \mathrm{Cov}\left[\prod_{j \in \mathcal{S}} \frac{z_j - p_j}{p_j(1-p_j)} \prod_{j' \in \mathcal{S}'} z_{j'}, \prod_{k \in \mathcal{T}} \frac{z_k - p_k}{p_k(1-p_k)} \prod_{k' \in \mathcal{T}'} z_{k'} \right] \right|.$$

By plugging in our bounds from Lemmas 3 and 4, we can simplify this bound:

$$\mathrm{Var}[\widehat{\mathrm{TTE}}] \leq \frac{1}{n^2} \sum_{i=1}^{n} \sum_{i' \in \mathcal{M}_i} \sum_{\mathcal{S}' \in \mathcal{S}_i^\beta} |c_{i,\mathcal{S}'}| \sum_{\mathcal{T}' \in \mathcal{S}_{i'}^\beta} |c_{i',\mathcal{T}'}| \sum_{\mathcal{S} \in \mathcal{S}_i^\beta} \sum_{\mathcal{T} \in \mathcal{S}_{i'}^\beta} \mathbb{I}(\mathcal{S} \triangle \mathcal{T} \subseteq \mathcal{S}' \cup \mathcal{T}') \cdot \left(\frac{1}{p(1-p)} \right)^{|\mathcal{S} \cap \mathcal{T}|}.$$

Via the change of variables $\mathcal{U} = \mathcal{S} \cap \mathcal{T}$, $\mathcal{S}'' = \mathcal{S} \backslash \mathcal{U}$, $\mathcal{T}'' = \mathcal{T} \backslash \mathcal{U}$, we may rewrite this

$$\begin{aligned}
&\frac{1}{n^2} \sum_{i=1}^{n} \sum_{i' \in \mathcal{M}_i} \sum_{\mathcal{S}' \in \mathcal{S}_i^\beta} |c_{i,\mathcal{S}'}| \sum_{\mathcal{T}' \in \mathcal{S}_{i'}^\beta} |c_{i',\mathcal{T}'}| \sum_{\mathcal{U} \in \mathcal{S}_i^\beta} \left(\frac{1}{p(1-p)} \right)^{|\mathcal{U}|} \#\left\{ (\mathcal{S}'', \mathcal{T}'') \in (\mathcal{S}' \cup \mathcal{T}')^2 : \begin{array}{c} \mathcal{S}'' \cap \mathcal{T}'' = \emptyset \\ |\mathcal{S}''|, |\mathcal{T}''| \leq \beta - |\mathcal{U}| \end{array} \right\} \\
&\leq \frac{1}{n^2} \sum_{i=1}^{n} \sum_{i' \in \mathcal{M}_i} \sum_{\mathcal{S}' \in \mathcal{S}_i^\beta} |c_{i,\mathcal{S}'}| \sum_{\mathcal{T}' \in \mathcal{S}_{i'}^\beta} |c_{i',\mathcal{T}'}| \sum_{\mathcal{U} \in \mathcal{S}_i^\beta} \left(\frac{1}{p(1-p)} \right)^{|\mathcal{U}|} (2\beta)^{2\beta - 2|\mathcal{U}|} \\
&\leq \frac{1}{n^2} (2\beta)^{2\beta} \sum_{i=1}^{n} \sum_{i' \in \mathcal{M}_i} \sum_{\mathcal{S}' \in \mathcal{S}_i^\beta} |c_{i,\mathcal{S}'}| \sum_{\mathcal{T}' \in \mathcal{S}_{i'}^\beta} |c_{i',\mathcal{T}'}| \sum_{\mathcal{U} \in \mathcal{S}_i^\beta} \left(\frac{1}{4\beta^2 p(1-p)} \right)^{|\mathcal{U}|} \\
&\leq \frac{d_{\text{in}} d_{\text{out}} Y_{\max}{}^2}{n} (2\beta)^{2\beta} \sum_{k=0}^{\beta} \binom{d_{\text{in}}}{k} \left(\frac{1}{4\beta^2 p(1-p)} \right)^{k} \\
&\leq \frac{d_{\text{in}} d_{\text{out}} Y_{\max}{}^2}{n} \left(\frac{e d_{\text{in}}}{\beta} \cdot \max\left(4\beta^2, \frac{1}{p(1-p)} \right) \right)^{\beta}.
\end{aligned}$$

Here, the final inequality makes use of the bound $\sum_{k=0}^{\beta} \binom{d}{k} \leq \left(\frac{ed}{\beta} \right)^{\beta}$. Note that when $\beta = 1$, this bound simplifies to

$$\frac{e \; d_{\text{in}}^2 \; d_{\text{out}} \; Y_{\max}{}^2}{np(1-p)}.$$

C Proof of Theorem 2

We use a variation of LeCam's method, which allows us to recast the hardness of TTE estimation through the lens of hypothesis testing. We consider the following setting.

– The causal network is a d-regular directed graph (so $d_{\text{in}} = d_{\text{out}} = d$ for all nodes) on n nodes.

– The coefficients $\{c_{i,\mathcal{S}}\}$ are drawn from one of two possible Gaussian distributions, Γ_0 and Γ_1. The coefficients are mutually independent under both distributions with marginal probabilities

$$\Gamma_0 : c_{i,\mathcal{S}} \sim \begin{cases} 0 & |\mathcal{S}| < \beta, \\ N\left(\delta\left(\frac{d}{\beta}\right)^{-1}, \left(\frac{d}{\beta}\right)^{-1}\right) & |\mathcal{S}| = \beta, \end{cases} \quad \Gamma_1 : \quad c_{i,\mathcal{S}} \sim \begin{cases} 0 & |\mathcal{S}| < \beta, \\ N\left(-\delta\left(\frac{d}{\beta}\right)^{-1}, \left(\frac{d}{\beta}\right)^{-1}\right) & |\mathcal{S}| = \beta, \end{cases}$$

where δ is a parameter that we will fix later.

– All units have a uniform treatment probability p.

By using the mutual independence assumption, we see that under Γ_0, each $Y_i(\mathbf{1}) \sim N(\delta, 1)$ and $\mathrm{TTE} \sim N\left(\delta, \frac{1}{n}\right)$. Under Γ_1, each $Y_i(\mathbf{1}) \sim N(-\delta, 1)$ and $\mathrm{TTE} \sim N\left(-\delta, \frac{1}{n}\right)$.

We wish to compute a lower bound on the mean squared error of any estimator for TTE in this setting. To begin this calculation, we have

$$\begin{aligned} \inf_{\widehat{\mathrm{TTE}}} \sup_{\mathbf{c}} \mathbb{E}_{\mathbf{z}}[(\widehat{\mathrm{TTE}} - \mathrm{TTE})^2 \mid \mathbf{c}] &\geq \inf_{\widehat{\mathrm{TTE}}} \sup_{\mathbf{c}} \frac{\delta^2}{100} \Pr\left(|\widehat{\mathrm{TTE}} - \mathrm{TTE}| \geq \frac{\delta}{10} \mid \mathbf{c}\right) \\ &\geq \frac{\delta^2}{100} \inf_{\widehat{\mathrm{TTE}}} \max_{\Gamma \in \{\Gamma_0, \Gamma_1\}} \mathbb{E}_{\mathbf{c} \sim \Gamma}\left[\Pr\left(|\widehat{\mathrm{TTE}} - \mathrm{TTE}| \geq \frac{\delta}{10} \mid \mathbf{c}\right)\right]. \end{aligned} \tag{A1}$$

Here, the first inequality lower bounds the conditional expectation by $\frac{\delta^2}{100}$ whenever $|\widehat{\mathrm{TTE}} - \mathrm{TTE}| \geq \frac{\delta}{10}$ and by 0 whenever $|\widehat{\mathrm{TTE}} - \mathrm{TTE}| < \frac{\delta}{10}$. The second inequality replaces the supremum over all possible $\mathbf{c}$ with a maximum over two possible distributions over $\mathbf{c}$.

Now, consider designing a hypothesis test Ψ to distinguish these models, i.e., a test for $\mathbb{1}(\mathbb{E}_{\mathbf{c}}[\mathrm{TTE}] > 0)$. Each estimator $\widehat{\mathrm{TTE}}$ gives rise to a decision rule $\widehat{\Psi} = \mathbb{1}(\widehat{\mathrm{TTE}} > 0)$. If $\widehat{\Psi}$ is incorrect, then one of the following two scenarios must have occurred:

(1) $|\mathrm{TTE} - \mathbb{E}_{\mathbf{c}}[\mathrm{TTE}]| \geq \frac{9\delta}{10}$ by a Gaussian tail bound this has probability $< \exp\left(\frac{-81n\delta^2}{200}\right)$.

(2) $|\widehat{\mathrm{TTE}} - \mathrm{TTE}| \geq \frac{\delta}{10}$.

When neither of these conditions is true, then $|\widehat{\mathrm{TTE}} - \mathbb{E}_{\mathbf{c}}[\mathrm{TTE}]| < \delta$, so $\widehat{\mathrm{TTE}}$ and $\mathbb{E}_{\mathbf{c}}[\mathrm{TTE}]$ have the same sign. By applying a union bound, we have

$$\Pr(\widehat{\Psi} \neq \mathbb{1}(\mathbb{E}_{\mathbf{c}}[\mathrm{TTE}] > 0)) < \exp\left(\frac{-81n\delta^2}{200}\right) + \Pr\left(|\widehat{\mathrm{TTE}} - \mathrm{TTE}| \geq \frac{\delta}{10}\right).$$

By rearranging this inequality and plugging into (A1), we may continue the simplification

$$\begin{aligned} &\geq \frac{\delta^2}{100} \inf_{\widehat{\mathrm{TTE}}} \max_{\Gamma \in \{\Gamma_0, \Gamma_0\}} \mathbb{E}_{\mathbf{c} \sim \Gamma}[\Pr(\widehat{\Psi} \neq \mathbb{1}(\mathbb{E}[\mathrm{TTE}] > 0) \mid \mathbf{c})] - \frac{\delta^2}{100} \exp\left(\frac{-81n\delta^2}{200}\right) \\ &\geq \frac{\delta^2}{100} \inf_{\Psi} \max_{\Gamma \in \{\Gamma_0, \Gamma_0\}} \mathbb{E}_{\mathbf{c} \sim \Gamma}[\Pr(\Psi \neq \mathbb{1}(\mathbb{E}[\mathrm{TTE}] > 0) \mid \mathbf{c})] - \frac{\delta^2}{100} \exp\left(\frac{-81n\delta^2}{200}\right) \\ &\geq \frac{\delta^2}{200} \inf_{\Psi} \left(\mathbb{E}_{\mathbf{c} \sim \Gamma_0}[\Pr(\Psi \neq 0)] + \mathbb{E}_{\mathbf{c} \sim \Gamma_1}[\Pr(\Psi \neq 1)] \right) - \frac{\delta^2}{100} \exp\left(\frac{-81n\delta^2}{200}\right) \\ &= \frac{\delta^2}{200}(1 - \|P_0 - P_1\|_{\mathrm{TV}}) - \frac{\delta^2}{100} \exp\left(\frac{-81n\delta^2}{200}\right) \\ &\geq \frac{\delta^2}{200}\left(1 - \sqrt{1 - \exp(-D_{\mathrm{KL}}(P_0 \| P_1))}\right) - \frac{\delta^2}{100} \exp\left(\frac{-81n\delta^2}{200}\right). \end{aligned}$$

Here, the second line follows because we have expanded the support of the infimum to all hypothesis tests, not just those that make use of an estimator $\widehat{\mathrm{TTE}}$. The third line lower bounds the maximum over the distributions by an average. The P_i in the fourth line represent the joint distribution over $(\mathbf{z}, \mathbf{c})$ when $\mathbf{c}$ is drawn from Γ_i. The last line is an application of the Bretagnolle-Huber inequality.

Next, we derive an upper bound for this KL-divergence. By applying the definition, we have

$$D_{\mathrm{KL}}(P_0||P_1) = \mathbb{E}_{P_0}\left[\log\left(\frac{P_0(\mathbf{Y},\mathbf{z})}{P_1(\mathbf{Y},\mathbf{z})}\right)\right] = \mathbb{E}_{P_0}\left[\log\left(\frac{P_0(\mathbf{Y}|\mathbf{z})}{P_1(\mathbf{Y}|\mathbf{z})}\right)\right].$$

Here, the second equality uses the fact that the treatment assignments are independent from the random model coefficients. Now, conditioned on the treatment assignment $\mathbf{z}$, the outcomes $\mathbf{Y}$ are distributed according to a Gaussian with independent coordinates. If we let $\mathcal{T}(\mathbf{z}) = \{i \in [n] : z_i = 1\}$ denote the set of treated individuals under $\mathbf{z}$, then the marginal distribution of $Y_i(\mathbf{z})$ conditioned on $\mathbf{z}$ is

$$Y_i(\mathbf{z}) \sim N\left(\pm\binom{|\mathcal{N}_i \cap \mathcal{T}(\mathbf{z})|}{\beta}\binom{d}{\beta}^{-1}\delta, \binom{|\mathcal{N}_i \cap \mathcal{T}|}{\beta}\binom{d}{\beta}^{-1}\right),$$

where the positive expectation comes from P_0 and the negative expectation comes from P_1. By plugging in the density of the Gaussian into our KL-divergence formula, we find that

$$\begin{aligned}
D_{\mathrm{KL}}(P_0||P_1) &= \mathbb{E}_{P_0}\left[\frac{-1}{2}\binom{d}{\beta}\sum_{i=1}^{n}\binom{|\mathcal{N}_i \cap \mathcal{T}(\mathbf{z})|}{\beta}^{-1}\left[\left(Y_i(\mathbf{z}) - \binom{|\mathcal{N}_i \cap \mathcal{T}(\mathbf{z})|}{\beta}\binom{d}{\beta}^{-1}\delta\right)^2 - \left(Y_i(\mathbf{z}) + \binom{|\mathcal{N}_i \cap \mathcal{T}(\mathbf{z})|}{\beta}\binom{d}{\beta}^{-1}\delta\right)^2\right]\right] \\
&= 2\delta\sum_{i=1}^{n}\mathbb{E}_{P_0}[Y_i(\mathbf{z})] \\
&= 2\delta^2\sum_{i=1}^{n}\binom{d}{\beta}^{-1}\mathbb{E}_{\mathbf{z}}\left[\binom{|\mathcal{N}_i \cap \mathcal{T}(\mathbf{z})|}{\beta}\right] \\
&= 2\delta^2\sum_{i=1}^{n}\binom{d}{\beta}^{-1}\sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i \\ \mathcal{S}=\beta}}\mathbb{E}_{\mathbf{z}}[\mathbb{I}(\mathcal{S} \subseteq \mathcal{T}(\mathbf{z}))] \\
&= 2n\delta^2 p^\beta
\end{aligned}$$

By plugging into our earlier results, we find that

$$\inf_{\widehat{\mathrm{TTE}}}\sup_{\mathbf{c}}\mathbb{E}_{\mathbf{z}}[(\widehat{\mathrm{TTE}} - \mathrm{TTE})^2|\mathbf{c}] \geq \frac{\delta^2}{200}\left(1 - \sqrt{1 - \exp(-2n\delta^2 p^\beta)} - 2\exp\left(\frac{-81n\delta^2}{200}\right)\right).$$

By taking $\delta^2 = \frac{8}{3np^\beta}$, we obtain the upper bound

$$\frac{1}{75np^\beta}\left(1 - \sqrt{1 - \exp\left(\frac{-16}{3}\right)} - 2\exp\left(\frac{-27}{25p^\beta}\right)\right).$$

This is $\Omega\left(\frac{1}{np^\beta}\right)$ as long as $p^\beta < 0.16$.

D Proof of Theorem 3

Proof of Theorem 3. We apply Theorem 3.6 from ref. [53] to the following defined random variables,

$$X_i \coloneqq \frac{1}{n}(Y_i w_i(\mathbf{z}) - \mathbb{E}[Y_i w_i(\mathbf{z})]), \quad v^2 \coloneqq \mathrm{Var}\left(\sum_{i\in[n]} X_i\right), \quad \text{and} \quad W \coloneqq \frac{1}{v}\sum_{i\in[n]} X_i,$$

where $w_i(\mathbf{z}) = \sum_{|\mathcal{S}|\leq\beta}^{\mathcal{S}\subseteq\mathcal{N}_i} g(\mathcal{S})\prod_{j\in\mathcal{S}}\left(\frac{z_j}{p_j} - \frac{1-z_j}{1-p_j}\right)$. By construction, it follows that

$$W = \frac{1}{nv}\sum_{i\in[n]}(Y_i w_i(\mathbf{z}) - \mathbb{E}[Y_i w_i(\mathbf{z})]) = \frac{1}{v}(\widehat{\mathrm{TTE}} - \mathrm{TTE}),$$

such that $\widehat{\mathrm{TTE}} = Wv + \mathrm{TTE}$. The proof follows from verifying the conditions used in Theorem 3.6 of [53], computing appropriate bounds for the moments of X_i, and using the fact that $v^2 = \Omega(1/n)$ by Assumption 3 and the variance bound in Theorem 1.

First, we note that $\mathbb{E}[X_i] = 0$ by construction. To upper bound $\mathbb{E}[X_i^4]$, note that that for all i, we have

$$|w_i(\mathbf{z})| = \left|\sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i \\ |\mathcal{S}|\leq\beta}} g(\mathcal{S})\prod_{j\in\mathcal{S}}\left(\frac{z_j}{p_j} - \frac{1-z_j}{1-p_j}\right)\right| \leq \left|\sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i \\ |\mathcal{S}|\leq\beta}} \frac{1}{p^{|\mathcal{S}|}}\right| \leq \left(\frac{d}{p}\right)^{\beta}. \tag{A1}$$

The second inequality in (A1) follows from Lemma 3, which upper bounds $|g(\mathcal{S})| \leq 1$. Furthermore, we use the assumption that for all i, $p_i \in [p, 1-p]$, and hence, $|(z_j/p_j - (1-z_j)/(1-p_j))| \leq 1/p$. The final inequality in (A1) follows from the fact that $|\mathcal{S}| \leq \beta$, and the number of subsets of $\mathcal{N}_i$ for any i is bounded above by d^{β}.

By recognizing that $\mathbb{E}[Y_i w_i(\mathbf{z})] = \sum_{1\leq|\mathcal{S}'|\leq\beta}^{\mathcal{S}'\subseteq\mathcal{N}_i} c_{i,\mathcal{S}'} \leq Y_{\max}$, we obtain a bound $|X_i| \leq \frac{1}{n}\left(Y_{\max}\left(\frac{d}{p}\right)^{\beta} + Y_{\max}\right)$ using the triangle inequality and the bound on $|w_i(\mathbf{z})|$ from (A1). From this, we can bound

$$\mathbb{E}[X_i^4] \leq \frac{1}{n^4}\mathbb{E}\left[\left(Y_{\max}\left(\frac{d}{p}\right)^{\beta} + Y_{\max}\right)^4\right] = \frac{Y_{\max}^4}{n^4}\left[\left(\frac{d}{p}\right)^{4\beta} + 4\left(\frac{d}{p}\right)^{3\beta} + 6\left(\frac{d}{p}\right)^{2\beta} + 4\left(\frac{d}{p}\right)^{\beta} + 1\right]. \tag{A2}$$

Since $d/p > 1$ and $\beta \geq 1$, we can then bound

$$\mathbb{E}[X_i^4] = O\left(\frac{1}{n^4}\left(\frac{d}{p}\right)^{4\beta} Y_{\max}^4\right). \tag{A3}$$

In the same way, we can bound

$$\mathbb{E}[|X_i|^3] \leq \frac{1}{n^3}\mathbb{E}\left[\left(Y_{\max}\left(\frac{d}{p}\right)^{\beta} + Y_{\max}\right)^3\right] = \frac{Y_{\max}^3}{n^3}\left[\left(\frac{d}{p}\right)^{3\beta} + 3\left(\frac{d}{p}\right)^{2\beta} + 3\left(\frac{d}{p}\right)^{\beta} + 1\right].$$

Since $d/p > 1$ and $\beta \geq 1$, we obtain the bound

$$\mathbb{E}[|X_i|^3] = O\left(\frac{1}{n^3}\left(\frac{d}{p}\right)^{3\beta} Y_{\max}^3\right). \tag{A4}$$

Let $\mathcal{M}_i$ denote the set of individuals i' such that $\mathcal{N}_i \cap \mathcal{N}_{i'} \neq \emptyset$, i.e., all individuals i' that share an in-neighbors with individual i. The set $\mathcal{M}_i$ characterizes the *dependency neighborhood* of i, as X_i and X_j are dependent if and only if there is a shared neighbor k such that X_i and X_j both depend on z_k. It follows that the maximum size of any dependency neighborhood $D = \max_{i\in[n]}|\mathcal{M}_i| \leq d_{\mathrm{in}}d_{\mathrm{out}} \leq d^2$. Then, by plugging in the bounds for D, $\mathbb{E}[|X_i|^3]$ and $\mathbb{E}[X_i^4]$ into Theorem 3.6 of ref. [53] results in the following bound

$$d_{\mathrm{W}}(W, Z) \leq O\left(\frac{d^4}{v^3}\frac{Y_{\max}^3 d^{3\beta}}{n^2 p^{3\beta}} + \frac{d^3}{v^2}\sqrt{\frac{Y_{\max}^4 d^{4\beta}}{n^3 p^{4\beta}}}\right), \tag{A5}$$

where recall that Z is a standard normal random variable. Since Assumption 3 implies that $v^2 \geq O(1/n)$, it follows that

$$d_{\mathrm{W}}(W, Z) = O\left(\frac{Y_{\max}^3 d^{3\beta+4}}{n^{1/2}p^{3\beta}} + \frac{Y_{\max}^2 d^{2\beta+3}}{n^{1/2}p^{2\beta}}\right).$$

By boundedness of $Y_{\max}$, d, β, and as $p = \omega(n^{-1/4\beta})$, the Wasserstein distance between W and $Z \sim N(0,1)$ goes to 0 as $n \to \infty$. As $\widehat{\mathrm{TTE}} = W\nu + \mathrm{TTE}$, it follows that the distribution of $\widehat{\mathrm{TTE}}$ converges to a normal with mean TTE and variance ν^2. □

E Other estimands

E.1 Average treatment effect

The *average treatment effect* (ATE) measures the average effect that one's own treatment has on their outcome (assuming no one else is treated).

$$\mathrm{ATE} \coloneqq \frac{1}{n}\sum_{i=1}^{n}(Y_i(\mathbf{e_i}) - Y_i(\mathbf{0})) = \frac{1}{n}\sum_{i=1}^{n} c_{i,\{i\}}.$$

The estimator for ATE takes the form

$$\widehat{\mathrm{ATE}} = \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}) \sum_{\mathcal{S}\in\mathcal{S}_i^\beta} (A_i)_{\mathcal{S},\{i\}} \prod_{j\in\mathcal{S}} z_j,$$

where A_i is the inverse of $\mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\top]$ as defined in Section 1. By plugging in the explicit form of these matrix entries, we have:

$$\begin{aligned}
\widehat{\mathrm{ATE}} &= \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}) \sum_{\mathcal{S}\in\mathcal{S}_i^\beta} \frac{-1}{p_i} \prod_{j\in\mathcal{S}} \frac{-z_j}{p_j} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta \\ (\mathcal{S}\cup\{i\})\subseteq\mathcal{U}}} \prod_{\ell\in\mathcal{U}} \frac{p_\ell}{1-p_\ell} \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}) \cdot \frac{-1}{p_i} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta \\ i\in\mathcal{U}}} \prod_{\ell\in\mathcal{U}} \frac{p_\ell}{1-p_\ell} \sum_{\mathcal{S}\subseteq\mathcal{U}} \prod_{j\in\mathcal{S}} \frac{-z_j}{p_j} \quad \text{(reverse inner sums)} \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}) \cdot \frac{-1}{p_i} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta \\ i\in\mathcal{U}}} \prod_{\ell\in\mathcal{U}} \frac{p_\ell}{1-p_\ell} \prod_{j\in\mathcal{U}} \frac{p_j - z_j}{p_j} \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}) \cdot \frac{-1}{p_i} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta \\ i\in\mathcal{U}}} \prod_{j\in\mathcal{U}} \frac{p_j - z_j}{1-p_j}.
\end{aligned}$$

E.2 Conditional average treatment effect

Given any subdemographic of the population $\mathcal{D} \subseteq [n]$, the *conditional average treatment effects* (CATE) of $\mathcal{D}$ is the average effect to an individual of the demographic that is treated in isolation

$$\mathrm{CATE}(\mathcal{D}) \coloneqq \frac{1}{|\mathcal{D}|}\sum_{i\in\mathcal{D}} (Y_i(\mathbf{e_i}) - Y_i(\mathbf{0})) = \frac{1}{|\mathcal{D}|}\sum_{i\in\mathcal{D}} c_{i,\{i\}}.$$

By the same calculation as mentioned earlier, we estimate the conditional average treatment effect:

$$\widehat{\mathrm{CATE}(\mathcal{D})} = \frac{1}{|\mathcal{D}|}\sum_{i\in\mathcal{D}} \widehat{c_{i,\{i\}}} = \frac{1}{|\mathcal{D}|}\sum_{i\in\mathcal{D}} Y_i(\mathbf{z}) \cdot \frac{-1}{p_i} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^\beta \\ i\in\mathcal{U}}} \prod_{j\in\mathcal{U}} \frac{p_j - z_j}{1-p_j}.$$

E.3 Size-dependent treatment effects

Although not standard in the literature, as its significance is largely brought about by the low-degree polynomial structure of our potential outcomes model, another family of causal estimands can be used to understand the magnitude of the treatment effects that individuals experience as a result of different sized subsets of their neighborhood being treated. We define the α-treatment effect,

$$\mathrm{TE}(\alpha) \coloneqq \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_i\\|\mathcal{S}'|=\alpha}} c_{i,\mathcal{S}'},$$

for some parameter $\alpha \le \beta$. That is, the α-treatment effect measures only those effects that subsets $\mathcal{S}'$ of size α have on the outcome of each individual and averages the cumulative effect over the individuals. For example, even though the polynomial degree of a model could be large, the causal effects associated to higher order interactions could be small such that the potential outcomes could be well approximated with a linear model ($\beta = 1$). In such an event, one would expect that TE(1) would be close to TTE, and TE(α) would be significantly smaller in magnitude for $\alpha > 1$. As TE(α) is again a linear combination of the model parameters, we can obtain an unbiased estimate using the same framework as mentioned earlier. In this case, the estimator takes the following form:

$$\begin{aligned}
\widehat{\mathrm{TE}(\alpha)} &= \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}) \sum_{\mathcal{S}\in\mathcal{S}_i^{\beta}} \sum_{\substack{\mathcal{T}\subseteq\mathcal{N}_i\\|\mathcal{T}|=\alpha}} (A_i)_{\mathcal{S},\mathcal{T}} \prod_{j\in\mathcal{S}} z_j \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}) \sum_{\mathcal{S}\in\mathcal{S}_i^{\beta}} \sum_{\substack{\mathcal{T}\subseteq\mathcal{N}_i\\|\mathcal{T}|=\alpha}} \prod_{j\in\mathcal{S}} \frac{-z_j}{p_j} \prod_{k\in\mathcal{T}} \frac{-1}{p_k} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^{\beta}\\(\mathcal{S}\cup\mathcal{T})\subseteq\mathcal{U}}} \prod_{\ell\in\mathcal{U}} \frac{p_\ell}{1-p_\ell} \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}) \sum_{\substack{\mathcal{T}\subseteq\mathcal{N}_i\\|\mathcal{T}|=\alpha}} \prod_{k\in\mathcal{T}} \frac{-1}{p_k} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^{\beta}\\\mathcal{T}\subseteq\mathcal{U}}} \prod_{\ell\in\mathcal{U}} \frac{p_\ell}{1-p_\ell} \sum_{\mathcal{S}\subseteq\mathcal{U}} \prod_{j\in\mathcal{S}} \frac{-z_j}{p_j} \qquad \text{(reorder sums)} \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}) \sum_{\substack{\mathcal{T}\subseteq\mathcal{N}_i\\|\mathcal{T}|=\alpha}} \prod_{k\in\mathcal{T}} \frac{-1}{p_k} \sum_{\substack{\mathcal{U}\in\mathcal{S}_i^{\beta}\\\mathcal{T}\subseteq\mathcal{U}}} \prod_{j\in\mathcal{U}} \frac{p_j - z_j}{1-p_j} \qquad \text{(distributivity)}.
\end{aligned}$$