A/B testing under Interference with Partial Network Information

Shiv Shankar
UMass

Ritwik Sinha
Adobe Research

Yash Chandak
Stanford University

Saayan Mitra
Adobe

Madalina Fiterau
UMass

Abstract

A/B tests are often required to be conducted on subjects that might have social connections. For e.g., experiments on social media, or medical and social interventions to control the spread of an epidemic. In such settings, the SUTVA assumption for randomized-controlled trials is violated due to network interference, or spill-over effects, as treatments to group A can potentially also affect the control group B. When the underlying social network is known exactly, prior works have demonstrated how to conduct A/B tests adequately to estimate the global average treatment effect (GATE). However, in practice, it is often impossible to obtain knowledge about the exact underlying network. In this paper, we present UNITE: a novel estimator that relax this assumption and can identify GATE while only relying on knowledge of the *superset* of neighbors for any subject in the graph. Through theoretical analysis and extensive experiments, we show that the proposed approach performs better in comparison to standard estimators.

1 INTRODUCTION

A/B tests, a form of randomized control trials, are the gold-standard in evaluating the impact of interventions, whether it be a new policy (Papadogeorgou et al., 2020) or a medical treatment (Antman et al., 1992), or experimentation in the digital world (Siroker and Koomen, 2015). A/B tests allow estimation of the treatment effect by ensuring that treatment (group A) and control (group B) assignments are made independently of other variables, including potentially unknown ones. The outcomes from the two groups are compared to determine the effect of treatment on a desired metric, such as health, clicks, or revenue.

However, in many scenarios, several assumptions required for the A/B testing protocol are violated. Particularly, in many large-scale experiments, there might exist an underlying social network through which the exposures to the treatment group might also affect the subjects in the control group. Consider the example of social network choosing a new algorithm to deploy for user content recommendation. Evaluating the effectiveness of a single algorithm can be challenging as any content (e.g., music, news, travel suggestions) recommended to a user in the treatment group might get shared with, and thus affect, another user in the control group if the two users have a social connection (Brennan et al., 2022; Pouget-Abadie et al., 2017), (Wong, 2020; Kusner et al., 2016). If the social network wants to evaluate the effectiveness of shifting to the newer algorithm, it wants to evaluate the GATE effect (where all units are provided the new treatment). However, if one ignores the effect of the these interactions, we might underestimate the effect of a new algorithm, as users getting recommendations from the older algorithm are still getting exposed to the newer algorithm indirectly. Other situations where this problem can manifest itself are:

(A) Privacy protection: To protect user's privacy, digital services may not be able to use cookies or other trackers for user identification (Shankar et al., 2023b). This can be problematic as users often access a digital service through multiple devices. For instance, if a device is assigned to be in the treatment group and another device in the control group, then the outcomes on those devices may not be independent if those devices are used by a common user.

(B) Epidemic control: Herd immunity is a phenomenon, whereby the virtue of a sufficiently large vaccinated/treated group the entire community gains immunity against a disease. Thus, the outcomes of the treated might be similar to the outcomes of the control, thereby leading to bias in the treatment effect estimate, if the community effect is not adequately accounted for (Randolph and Barreiro, 2020; Fine, 1993).

This phenomenon of treatment to one unit affecting outcomes for other units is known as interference (Hudgens and Halloran, 2008; LeSage and Pace, 2009). While interference-aware methods for treatment effect estimation exist, they often require knowledge of the interference structure (Ogburn et al., 2017; Leung, 2020). However, in practice, it can be challenging or impossible to obtain the exact network structure for post-experimentation analysis using these interference-aware methods as well. Thus the scenarios mentioned above introduce a new challenge, necessitating the estimation of treatment effects from an *uncertain* or a *partial view* of interference structure.

Contributions. We focus on the challenge of treatment effect estimation in the presence of interference. We consider the setting where there is an interaction graph among subjects, and interference emanates from 'neighboring' subjects. Compared to prior work, our focus is on scenarios where the exact neighborhood information is unavailable.

Our key contribution is to develop a method that identifies the Global Average Treatment Effect (GATE) using access to only the knowledge of a *superset of these interfering units*, which is relatively easier to obtain in practice (e.g., using geographic locality). Furthermore our estimator's variance matches the mini-max optimal lower bound on MSE (Ugander et al., 2013; Yu et al., 2022) suggesting the theoretical efficiency of our estimate. Our paper not only establishes the theoretical validity of our estimator but also substantiates its practical efficacy through extensive experiments on both simulated and real-world datasets.

2 RELATED WORK

Network interference is a well studied topic in causal inference literature, with a variety of methods proposed for the problem. These often use varied set of assumptions about the exposure function (Aronow et al., 2017; Auerbach and Tabord-Meehan, 2021; Li et al., 2021; Viviano, 2020) and interference neighborhoods (Bargagli-Stoffi et al., 2020; Bhattacharya et al., 2020; Sussman and Airoldi, 2017a; Ugander et al., 2013). Similar to these proposals, our approach also relies on imposing some structure to the form of network interference. A common assumption is that the network effects is linear with respect to a known functional of the neighbour treatments(Basse and Airoldi, 2018; Cai et al., 2015; Chin, 2019; Gui et al., 2015; Parker et al., 2017; Toulis and Kao, 2013).

Another common assumption is an exposure represented as the (weighted) proportion of either the neighboring units that have received treatment (Eck-

les et al., 2017; Toulis and Kao, 2013), or the count of neighboring units that have undergone treatment (Ugander et al., 2013). Frequently, a normalized exposure assumption is also applied, ensuring that for all units the net exposure is in $[-1, 1]$, where the extremes of $-1/1$ represent the unit being in control/ treatment group respectively (Leung, 2016). Such settings within the context of bipartite interference have been explored by researchers such as Pouget-Abadie et al. (2017), Pouget-Abadie et al. (2019), and Brennan et al. (2022). Furthermore, when the exposure model is known, various cluster randomized designs have been proposed for variance reduction in the estimate (Eckles et al., 2017; Gui et al., 2015; Pouget-Abadie et al., 2019). The key limitation of these methods is that one needs both the exact network structure and exposure model to compute the correct neighborhood statistics.

Recently, some methods have been proposed based on multiple measurements which can address interference without knowing the network (Shankar et al., 2023b; Cortez et al., 2022; Yu et al., 2022). These methods use a multiple-trial approach for estimating treatment effect. Assuming stationarity i.e. that the outcomes do not vary between the trials, multiple trials simplify GTE estimation by providing access to both the factual and counterfactual outcome. However, under non-stationarity (e.g. seasonal effects) their estimator is not valid. Moreover, such a model is unrealistic for our motivating use case of continuous optimization. Finally, in the more general setting, conducting multiple trials itself is fundamentally impossible (Shankar et al., 2023a). As such, we want to develop a method which can work with observational data from a single RCT.

Other works on treatment effect estimation under uncertain or mis-measured data include those focused on sensitivity analysis (Liu et al., 2013; Richardson et al., 2014; Veitch and Zaveri, 2020; Dorie et al., 2016) or on trying to identify confounders (Ranganath and Perotte, 2018; D'Amour, 2019; Wang and Blei, 2019; Miao et al., 2020). Yadlowsky et al. (2018) propose a loss-minimization approach that quantifies bounds on the conditional average treatment effect (CATE). They focus on analysis of errors in outcomes, confounder models, etc. but not on error in the interference graph.

3 UNITE

In this section, we first formalize the treatment effect estimation problem with uncertainty about the interference graph. We then present the assumptions that underlie our method. Finally, we describe our estimation method called *UNITE* (Uncertain Network
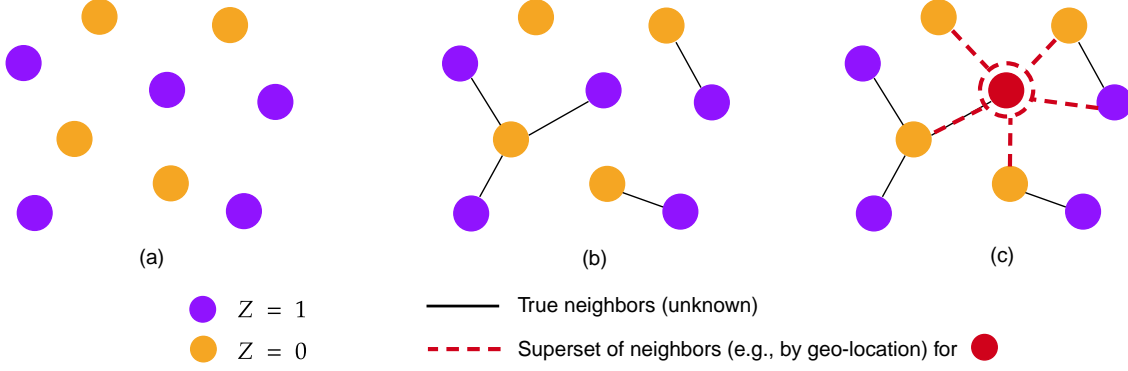
Figure 1: $Z = 1$ denotes to units in the treatment group and $Z = 0$ denotes units in the control group. **(a)** Standard A/B testing where there is no interaction between the treatment and the control units. **(b)** Network interference due to (unknown) interaction between the units. **(c)** We do not assume access to the exact underlying interaction graph. Instead, we consider a practically feasible assumption where only a *superset* of neighbors for any node is available.

Interference Aware Treatment-effect Estimator), and present theoretical results about the method.

3.1 Notation and Assumptions

Let there be n units in the population, $\boldsymbol{Z}$ be the treatment assignment vector of the entire population, $\mathcal{Z}$ denote the set of possible treatment assignments, e.g., for binary treatments $\mathcal{Z} = \{0, 1\}^n$. We use the Neyman potential outcome framework (Neyman, 1923; Rubin, 1974), and let $Y_i(\boldsymbol{z})$ denote the potential outcome of the i-th for $\boldsymbol{z} \in \mathcal{Z}$.

Observations are made only at the unit level and are denoted as Y_i for unit i. Note that multiple units might have a common factor influencing them. For example, if the units are users, their choice can be influenced not only by their experience on a website, but also by the experience of their social circle. Similarly, for a user using multiple devices, their behaviour at one device can be influenced by the treatment at the other device. As such the potential outcome at unit i need not depend only on its own treatment assignment but also on other treatments allocated to its neihgbours. This is a violation of the SUTVA assumption (Cox, 1958; Hudgens and Halloran, 2008); and is commonly called interference or spillover.

The dependence between unit level outcomes, can be encoded with its adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, with $A_{ij} = 1$ only if treatment at unit j can influence outcomes at unit i (e.g. if units j is a friend of unit i or if i and j are used by the same user). Let $\mathcal{N}_i = \{j : A_{ij} = 1\}$ be the set of *neighbors* of device i, where by convention we include the self node i.e. $A_{ii} = 1$. We assume that the outcomes depend on the total treatments received by a unit through the graph.

This implies that the interference at a device is limited to its immediate neighbours in the graph (i.e. SUTVA holds at the network level).

$$\text{If } z_i = z_i' \text{ and } z_j = z_j' \ \forall j \in \mathcal{N}_i, \text{ then } Y_i(\boldsymbol{z}) = Y_i(\boldsymbol{z}').$$

The desired causal effect (GATE) is the mean difference between the outcomes when $\boldsymbol{z} = \vec{1}$ (i.e., $\forall i, z_i = 1$) and when $\boldsymbol{z} = \vec{0}$ (i.e., $\forall i, z_i = 0$). Under the aforementioned notations, GATE is given by:

$$\tau(\vec{1}, \vec{0}) = \frac{1}{n} \sum_{i=1}^{n} Y_i(\vec{1}) - \frac{1}{n} \sum_{i=1}^{n} Y_i(\vec{0}) \qquad (1)$$

To estimate $\tau(\vec{1}, \vec{0})$, we consider the following assumptions, that can broadly categorized under the following three categories.

(I) Partial Graph: If the true graph is known, prior works have provided an estimate of $\tau(\vec{1}, \vec{0})$ (Hudgens and Halloran, 2008; Halloran and Hudgens, 2016). However, access to $\boldsymbol{A}$ is not possible in many settings. Therefore, we will assume that the **exact** graph $\boldsymbol{A}$ is **not** known. Instead, we assume access to a model $\mathcal{M}$ which provides constraints on $\boldsymbol{A}$. Specifically, we assume $\mathcal{M}$ can be queried for a node i to obtain a *superset* $\mathcal{M}_i \supseteq \mathcal{N}_i$ of the true neighborhood.

In our use case of experimenting with social graphs, a superset of neighbors can be obtained in various ways: if the user has given cookie permissions, or from some existing user model for identity linking, or even from something as basic as geo-location and ip addresses. This provides a significant practical advantage over the prior methods that necessitate knowledge of the *exact* neighborhood.

(II) Linear Additive Structure: Randomized experiments with interference (even with neighbourhood interference) can be difficult to analyze since the number of potential outcome functions for unit i may be $2^{|\mathcal{N}_i|}$. This *exponential* growth in number of neighbours $|\mathcal{N}_i|$ is problematic in comparison to the SUTVA case where there are only two outcome functions, irrespective of $|\mathcal{N}_i|$. Therefore, the literature on network interference often restricts the set of potential outcome functions. (Eckles et al., 2017; Toulis and Kao, 2013; Sussman and Airoldi, 2017a). We follow a similar approach and let the outcome model be

$$Y_i(\boldsymbol{z}) = c_i + \sum_{j \in \mathcal{N}_i} c_{i,j} \mathbb{I}[z_j = 1] + \varepsilon \qquad (\mathbf{A1})$$

where ε is an independent zero mean noise. Under this assumption, GATE is given by $\tau(\vec{1}, \vec{0}) = \frac{1}{n} \sum_i \sum_{j \in \mathcal{N}_i} c_{ij}$. We consider this linear model to explain the core idea of our estimator. In Section 3.4 we consider the generalized setting where Y_i can be a non-linear function.

(III) We also make the following assumptions that are standard for the treatment effect and causal interference (Pearl, 2009; Hudgens and Halloran, 2008; Spirtes, 2010):

Network Ignorability: $Y(\boldsymbol{z}) \perp\!\!\!\perp \boldsymbol{Z}$ $\forall \boldsymbol{z}$ **(A2)**

Positivity: $P(\boldsymbol{z}|\mathbf{X}) > 0$ $\forall \boldsymbol{z}$ **(A3)**

Consistency: $Y_i = Y_i(\boldsymbol{z})$ if $\boldsymbol{Z} = \boldsymbol{z}$ **(A4)**

Since we are in a A/B testing scenario, there do not exist any confounders. Moreover positivity is easily ensured by choosing a good randomization scheme. Hence the assumptions **A2-4** are naturally satisfied.

3.2 Estimation

One of the most popular estimators for causal inference is the Horvitz-Thompson (HT) estimator (Horvitz and Thompson, 1952). If the graph is known and when all treatment decisions are independent Bernoulli variables with probability p, an estimate τ_{HT} of GATE using HT estimator can be obtained as the following:

$$\tau_{\mathrm{HT}} = \frac{1}{n} \sum_i Y_i \left(\prod_{j \in \mathcal{N}_i} \frac{z_j}{p} - \prod_{j \in \mathcal{N}_i} \frac{(1-z_j)}{(1-p)} \right). \quad (2)$$

However, using τ_{HT} is practically infeasible because of the *exponentially* high variance. To observe the cause of high variance, consider that the graph $\boldsymbol{A}$ is a k-regular graph and $p = 0.5$. In such a case, the probability that any given node i is exposed to only treatment or control group nodes in the treatment or control group is $1/2^k$. This falls off rapidly with k, requiring large graphs for even modest k. For example,

with $k = 20$, we need a graph of the order of a million nodes for the HT estimate to even have a meaningful value. This problem of the HT-estimate is further compounded when the graph $\boldsymbol{A}$ is not exactly known.

This raises a natural question: *can the structure of the problem be exploited to reduce variance?*

In the following, we show how the structure in **A1** can be leveraged to simultaneously address both the challenges: a) high variance of τ_{HT} and b) incomplete knowledge of the graph $\boldsymbol{A}$. We will first present a basic estimator, which we would then improve via incorporating ideas from self-normalized estimators (Thomas and Brunskill, 2017) and doubly-robust estimation (Chernozhukov et al., 2018).

3.3 Variance Reduced Estimation

We first have a look at how assumption **A1** affects τ_{HT}. Substituting **A1** in 2, τ_{HT} can be expressed as

$$\frac{1}{n} \sum_i \left[c_i + \sum_{j \in \mathcal{N}_i} c_{i,j} \mathbb{I}[z_j = 1] \right] \left(\prod_{k \in \mathcal{N}_i} \frac{z_k}{p} - \prod_{k \in \mathcal{N}_i} \frac{(1-z_k)}{(1-p)} \right).$$

Now observe that as *allocation* at each unit is independent, for any functions g and h: $\mathbb{E}[h(z_i)g(z_j)] = \mathbb{E}[h(z_i)]\mathbb{E}[g(z_j)]$. Furthermore, as $\mathbb{E}[z_k/p] = \mathbb{E}[(1 - z_k)/(1-p)] = 1$, we can ignore all the ratio terms for $k \neq j$ (see Appendix D for a complete derivation). Therefore, τ_{HT} can be simplified as

$$\mathbb{E}[\tau_{\mathrm{HT}}] = \frac{1}{n} \sum_i \mathbb{E}\left[[c_i + \sum_{j \in \mathcal{N}_i} c_{i,j} \mathbb{I}[z_j = 1]] \left(\frac{z_j}{p} - \frac{(1-z_j)}{(1-p)} \right) \right],$$

which is a linear combination of in the terms z_j/p and $(1-z)/(1-p)$. However, while this avoids the root cause of high variance (the product of the ratios), this expression cannot be computed from only the graph and observed outcomes Y_i.

To resolve this problem, we will rewrite the earlier expression in terms of Y_i. Observe that since $z_j \perp\!\!\!\perp z_i \ \forall i \neq j$, we can add terms of the form $z_i \left(\frac{z_j}{p} - \frac{1-z_j}{1-p} \right)$ with $i \neq j$ without changing the expected value. Adding in such terms to include every node in $\mathcal{M}_i$, we get

$$\mathbb{E}[\tau_{\mathrm{HT}}] = \frac{1}{n} \sum_i \mathbb{E}\left[\left(c_i + \sum_{j \in \mathcal{N}_i} c_{i,j} \mathbb{I}[z_j = 1] \right) \left(\sum_{k \in \mathcal{M}_i} \frac{z_k}{p} - \frac{(1-z_k)}{(1-p)} \right) \right]$$

which motivates the following estimator:

$$\hat{\tau}_{\mathrm{Lin}} = \frac{1}{n} \sum_i Y_i \sum_{j \in \mathcal{M}_i} \left(\frac{z_j}{p} - \frac{(1-z_j)}{(1-p)} \right). \quad (3)$$

Theorem 1. *Under **A1-4**, and assuming $\mathcal{M}_i \supseteq \mathcal{N}_i$, $\hat{\tau}_{Lin}$ is an unbiased and consistent estimate of $\tau(\vec{1}, \vec{0})$, i.e., $\mathbb{E}[\hat{\tau}_{Lin}] = \tau(\vec{1}, \vec{0})$, and $\hat{\tau}_{Lin} \xrightarrow{a.s.} \tau(\vec{1}, \vec{0})$.*

Remark 1. *Observe that both 2 and 3 are linear combinations of outcomes Y_i but the weights are different. While τ_{HT} uses weights that involve products over the true neighbors $\mathcal{N}_i$ of a node, $\hat{\tau}_{Lin}$ use weights that are summed over a superset of neighbors $\mathcal{M}_i$. This allows a drastic reduction in variance.*

Remark 2. *$\hat{\tau}_{Lin}$ is unbiased under the structural assumption **A1**, but τ_{HT} is unbiased irrespective. Intuitively, variance is traded-off with the identifiability of τ using the structural form of interference.*

Self-Normalization While $\hat{\tau}_{\text{Lin}}$ can drastically reduce variance compared to τ_{HT}, it can still be subject to high variance when p is close to 0 in the inverse-propensity (z_i/p) terms. This fact is well known in the importance sampling (IS) literature where the inverse propensity is also known as 'importance ratios'. A common alternative there is to use the self-normalized (or weighted) IS that can further reduce the variance at the cost of bias (Tukey, 1956; Rubin, 1987).

Note that $\hat{\tau}_{\text{Lin}}$ bears resemblance to the general IS estimators, where in 3 the numerator in the ratios is the probability of treatments with the desired distribution (i.e., assign all units 0 or 1), and the denominator is the probability of the treatment under the sampling distribution. Using this insight, we can leverage self-normalization to further reduce the variance of $\hat{\tau}_{\text{Lin}}$.

Let $\rho_i = z_i/p$ and $\rho'_i = \frac{1-z_i}{1-p}$, then $\hat{\tau}_{\text{Lin}}$ can be rewritten with ρ, and ρ' as:

$$\hat{\tau}_{\text{Lin}} = \frac{1}{n} \sum_i Y_i \sum_{j \in \mathcal{M}_i} \left(\rho_j - \rho'_j \right).$$

Similar to self-normalized estimators, we can replace the importance weights ρ by their self-normalized values $\bar{\rho}_i$ and $\bar{\rho}'_i$, where

$$\bar{\rho}_i = \frac{\rho_i}{\frac{1}{n} \sum_j \frac{1}{|M_j|} \sum_{k \in M_j} \rho_k}, \quad \bar{\rho}'_i = \frac{\rho'_i}{\frac{1}{n} \sum_j \frac{1}{|M_j|} \sum_{k \in M_j} \rho'_k}. \quad (4)$$

The self-normalized estimator in our case is given by:

$$\hat{\tau}^1_{WIS} = \frac{1}{n} \sum_i Y_i \left(\sum_{k \in \mathcal{M}_i} \bar{\rho}_k - \bar{\rho}'_k \right).$$

Theorem 2. *Under **A1-4** and assuming $\mathcal{M}_i \supseteq \mathcal{N}_i$, $\hat{\tau}^1_{WIS}$ is a biased but consistent estimate of $\tau(\vec{1}, \vec{0})$, i.e., $\mathbb{E}[\hat{\tau}^1_{WIS}] \neq \tau(\vec{1}, \vec{0})$, and $\hat{\tau}^1_{WIS} \xrightarrow{a.s.} \tau(\vec{1}, \vec{0})$.*

Doubly Robust Estimate: Through self-normalization, $\hat{\tau}^1_{WIS}$ reduces the variance at the cost of bias. To avoid incurring this bias, but still reduce the variance of $\hat{\tau}_{\text{Lin}}$, we now propose a doubly robust estimator that leverages partial estimates of $\tau(\vec{1}, \vec{0})$.

There are a plethora of methods that make stronger assumptions about the interference or graph structure (e.g., knowledge of the exact graph) (Brennan et al., 2022; Eckles et al., 2017; Leung, 2020). These methods will typically provide biased but lower variance estimates when their assumptions do not hold. However, we can use our method to create a 'robustified' version of these existing estimators, which not only (a) side-steps the need to validate the assumptions such models make, but can also (b) potentially reduce the variance of our estimator.

Leveraging ideas from control-variate (Thomas and Brunskill, 2016) and doubly robust learning (Chernozhukov et al., 2018) literature, given a (partial/incorrect) estimate of the potential outcome functions $f_0(X_i) \approx Y_i(\vec{0})$ and $f_1(X_i) \approx Y_i(\vec{1})$, we propose the following estimator:

$$\hat{\tau}_{DR} = \frac{1}{n} \sum_i \Big[f_1(X_i) - f_0(X_i) + $$
$$(Y_i - f_1(X_i)) \sum_{\mathcal{M}_i} \rho_j - (Y_i - f_0(X_i)) \sum_{\mathcal{M}_i} \rho'_j \Big].$$

Estimator $\hat{\tau}_{DR}$ is beneficial as it may reduce variance using (partial) models f_0 and f_1, and remain unbiased irrespective of how inaccurate f_0 and f_1 are.

Theorem 3. *Under **A1-4** and assuming $\mathcal{M}_i \supseteq \mathcal{N}_i$, $\hat{\tau}_{DR}$ is an unbiased and consistent estimate of $\tau(\vec{1}, \vec{0})$, $\mathbb{E}[\hat{\tau}_{DR}] = \tau(\vec{1}, \vec{0})$, and $\hat{\tau}_{DR} \xrightarrow{a.s.} \tau(\vec{1}, \vec{0})$.*

3.4 Non-linear Structural Model

Motif Model In the earlier section, we considered a linear structural model from **A1**. However, in general, there may also exist interaction from subgraphs in the neighborhood of i. Graphs for such interaction structure are also known as "motif" models (Aittokallio and Schwikowski, 2006; Holland and Leinhardt, 1974), as "network motifs" are a common way to characterize all patterns of smaller network features among a set of nodes (see Figure 2 for examples). While this may seem restrictive, our assumption on the functional form of potential outcome is less stringent than that commonly used in literature (Sussman and Airoldi, 2017b; Brennan et al., 2022). Furthermore if the neighborhoods are small, the motif model can represent all possible functions (Yuan et al., 2021), and hence can also capture model heterogeneity in interference.

To model this setting, we consider outcomes to be a linear combination of *influences* from small-sized "network motifs",

$$Y_i(\boldsymbol{z}) = c_i + \sum_{S \in \mathcal{SN}_i^\beta(\boldsymbol{A})} c_{i,S} \prod_{j \in S} \mathbb{I}[z_j = 1] + \varepsilon, \quad (\textbf{A5})$$
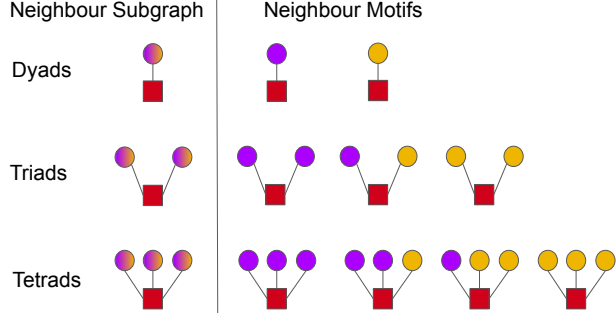
Figure 2: Examples of network motifs. Red square represent a node under consideration, and circles are its neighbours. the left column lists possible subgraphs of the neighbourhood of a node. The right hand side depicts different motifs corresponding to these subgraphs. The yellow node represents $Z = 0$ while purple represents $Z = 1$. A motif is activated when its nodes are assigned the corresponding treatment assignment

where $\mathcal{SN}_i(\boldsymbol{A})$ is a set of motifs of size up to β for node i. It can equivalently be considered as a collection of subsets of $\mathcal{N}_i$ such that no set in $\mathcal{SN}_i(\boldsymbol{A})$ has size $> \beta$.

Note that, if we consider only dyads, i.e. we set $\beta = 1$ and consider all such dyadic elements in $\mathcal{N}_i$ then **A5** is equivalent to **A1**. Considering $\beta > 1$ allows non-linear interaction between the nodes.

Estimator An important advantage of the estimators developed in the previous section is that they can be readily extended for the setting with non-linear interaction among the nodes. Specifically, let

$$\hat{\tau}^\beta = \frac{1}{n} \sum_{i=1}^n Y_i \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1 - p} \right)$$

Theorem 4. *Under **A2-5** and assuming $\mathcal{M}_i \supseteq \mathcal{N}_i$, $\mathbb{E}[\hat{\tau}^\beta] = \tau(\vec{1}, \vec{0})$, and $\hat{\tau}^\beta \xrightarrow{a.s.} \tau(\vec{1}, \vec{0})$.*

Self-Normalized and Doubly Robust: Once again, expressing $\hat{\tau}^\beta$ using $\rho_i = z_j/p$ and $\rho'_i = \frac{1-z_j}{1-p}$,

$$\hat{\tau}^\beta = \frac{1}{n} \sum_{i=1}^n Y_i \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} \left(\frac{z_j}{p} - 1 \right) - \prod_{j \in S} \left(-1 + \frac{1-z_j}{1-p} \right) \right)$$

$$= \frac{1}{n} \sum_{i=1}^n Y_i \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} (\rho_j - 1) - \prod_{j \in S} (-1 + \rho'_j) \right).$$

Let $\bar{\rho}_j$ and $\bar{\rho}'_j$ be the corresponding self normalized values of ρ and ρ', as in 4. The self-normalized estimate

for $\hat{\tau}^\beta$ is given by:

$$\hat{\tau}^\beta_{\text{WIS}} = \frac{1}{n} \sum_{i=1}^n Y_i \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} (\bar{\rho}_j - 1) - \prod_{j \in S} (-1 + \bar{\rho}'_j) \right)$$

Theorem 5. *Under **A2-5** and assuming $\mathcal{M}_i \supseteq \mathcal{N}_i$, $\mathbb{E}[\hat{\tau}^\beta_{WIS}] \neq \tau(\vec{1}, \vec{0})$, and $\hat{\tau}^\beta_{WIS} \xrightarrow{a.s.} \tau(\vec{1}, \vec{0})$*

Similarly, if we have access to a model for the counterfactual outcomes $f_0(X_i) \approx Y_i(\vec{0})$ and $f_1(X_i) \approx Y_i(\vec{1})$, we can create a DR estimator as

$$\hat{\tau}^\beta_{DR} = \frac{1}{n} \sum_{i=1}^n \tilde{Y}_i \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} (\rho_j - 1) - \prod_{j \in S} (-1 + \rho'_j) \right)$$

$$+ \frac{1}{n} \sum_{i=1}^n (f_1(X_i) - f_0(X_i)) \qquad (5)$$

where $\tilde{Y}_i = Y_i - z_i(f_1(X_i) - (1 - z_i)f_0(X_i)$.

Theorem 6. *Under **A2-5** and assuming $\mathcal{M}_i \supseteq \mathcal{N}_i$, $\mathbb{E}[\hat{\tau}^\beta_{DR}] = \tau(\vec{1}, \vec{0})$ and $\hat{\tau}^\beta_{DR} \xrightarrow{a.s.} \tau(\vec{1}, \vec{0})$.*

Application to General Non-Linearity Finally, the motif model in **A5** paves the way for a general non-linear structural model. In the following, we will use the index j, k to refer to neighbors of node i. The dependence of their ranges and indices on i is implicit and suppressed for notational ease. Let the outcome for node i be determined by the function g_i. If g_i is an analytic function, using Taylor-polynomials,

$$Y_i(\boldsymbol{z}) = g_i(z_1, z_2, ..z_{|\mathcal{N}_i|})$$
$$= g_i(\vec{0}) + \sum_{z_j} (\partial_{z_j} g_i) z_j + \sum_{z_j, z_k} \left(\partial^2_{z_j, z_k} g_i \right) z_j z_k + O(\boldsymbol{z}^3)$$
$$\approx g_i(\vec{0}) + \sum_{z_j} (\partial_{z_j} g_i) z_j + \sum_{z_j, z_k} \left(\partial^2_{z_j, z_k} g_i \right) z_j z_k, \qquad (6)$$

which is of the same form as **A5** with $\beta = 2$. Therefore, for interference under a general non-linear structural model, the estimate from $\hat{\tau}^\beta$ with $\beta = 2$ may be biased but when g_i can be assumed to be sufficiently smooth, its bias can be bounded.

Theorem 7. *For the non-linear model 6, if g_i is $k + 1$ times differentiable and the $k + 1th$ derivative is bounded by C, then the absolute bias*

$$\left| \mathbb{E}[\hat{\tau}^\beta] - \tau(\vec{1}, \vec{0}) \right| \leq \frac{C \max(p, 1 - p)}{(k + 1)!}.$$

3.5 Statistical Inference

The results till now were focused with providing point-estimates of the treatment effect. However, in practice, one needs reasonable confidence intervals around

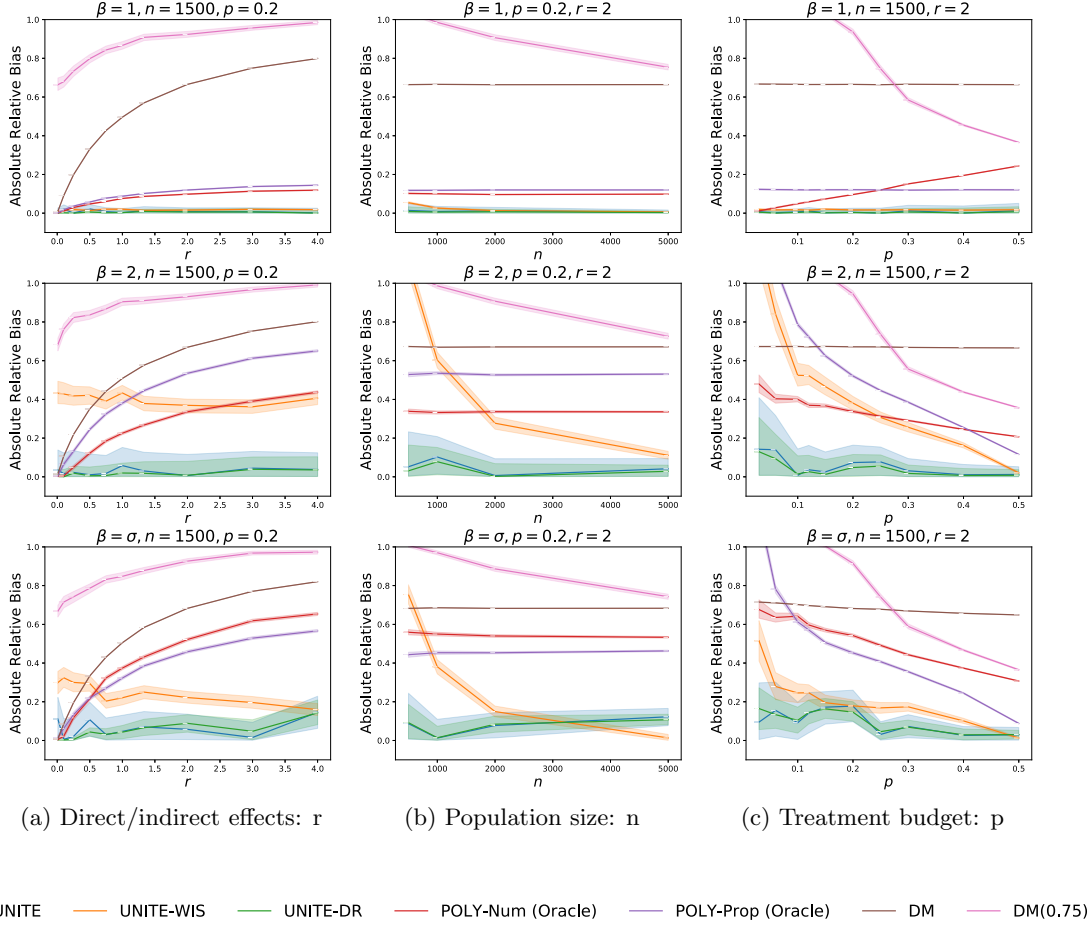(a) Direct/indirect effects: r   (b) Population size: n   (c) Treatment budget: p

Figure 3: Performance of GATE estimators under Bernoulli design on Erdos-Renyi networks. Y-axes represent the relative value of the absolute bias i.e. $\left|\frac{\mathbb{E}[\hat{\tau} - \tau(\vec{1}, \vec{0})]}{\tau}\right|$, with the shaded width corresponding to the experimental deviation. The rows correspond to linear, quadratic and sigmoid model for the potential outcomes. Columns correspond to different parameters being varied: (a) Strength of interference, where the x-axis corresponds to the average ratio of indirect to direct effects $r = \frac{1}{n}\sum_{i,j}\frac{|c_{ij}|}{|c_{ii}|}$. (b) Population size, where the x-axis corresponds to the number of nodes n (c) Treatment budget, where the x-axis corresponds to the probability of treatment 1 (p).

these estimates, to handle statistical uncertainty and perform hypothesis tests to verify assumptions. For this purpose, we first provide bounds for variance of the unbiased estimator $\hat{\tau}^\beta$.

Theorem 8. *Under **A2-5** and further assuming that* $\mathcal{M}_i \supseteq \mathcal{N}_i$, *if the ε noise have variance σ then,* $\mathrm{Var}[\hat{\tau}^\beta] \in O(1/n)$

Thus the UNITE estimator is consistent. We would like to highlight that the dependence is exponential in $d_\mathcal{M}^\beta$, where d is related to the degrees of nodes in the graph. Incorporating higher order motifs can lead to higher variance, highly dense graphs have high variance and overall consistency requires a growth constraint on the graph density/degree. For the exact dependence on various graph parameters, we refer the

reader to the Appendix. Similar bounds also hold for the DR and WIS estimators.

Next, we argue that this estimator is asymptotically normal. For this we rely on a classic result in generalized central limit theorems (Ross, 2011). Informally, for a set of n bounded random variables R_i, if the dependency graph is not too dense, then the variance normalized sum, $\sum R_i / \sqrt{(\mathrm{Var}(\sum R_i))}$ approach a gaussian distribution. We can clearly see that the estimator $\hat{\tau}^\beta$ (as well as other variants) can be written as a sum of such random variables. The dependence between the variables is represented by the neighbourhoods in $\mathcal{M}$. As such if $\mathcal{M}$ is not too dense, $\hat{\tau}^\beta$ is asymptotically normal. Further note that while we assume that the max-degree of $\mathcal{M}$ i.e $d_\mathcal{M}$ is constant, the

exact statement of the theorem allows sub-polynomial growth in the degree [1].

Combined with the variance bound, the normality results suggests a way to do statistical inference. Specifically, since the variance formula is an upper bound, we can construct conservative Wald-type intervals (Wasserman, 2006) without requiring the knowledge of the underlying graph. We should note however, that since the convergence is asymptotic, the use of the aforementioned variance for confidence intervals is only approximately valid. However, insights from Sussman and Airoldi (2017b) affirm the minimax optimality of this bound concerning its dependence on p, σ, and the max-degree of $\mathcal{M}$ δ. Consequently, these intervals maintain a robust reliability, yielding a level of precision considerably superior to that of the HT estimate (Aronow et al., 2017).

4   EXPERIMENTS

4.1   Synthetic Interference Graphs

In this section, we present synthetic experiments with interference on Erdos-Renyi graphs. The Erdos-Renyi (ER) model is commonly used for analyzing interaction networks in various experimental settings, particularly in the realm of social media (Seshadhri et al., 2012) and epidemic control (Kephart and White, 1992; Wang et al., 2003). In social media platforms, where connections form organically, ER graphs provide a reasonable simulation of how friendships, followerships, or interactions might evolve in an online community (Erdos et al., 1960). Additionally, in the context of epidemic control, ER graphs serve as valuable models for studying disease spread (Wang et al., 2003).

We simulate 50 different random Erdos-Renyi Graphs and run repeated experiments on these graphs with random treatment assignments. For these experiments, we provide an ablation study by varying the treatment probability, the strength of interference, and the size of the graphs to assess the efficacy of estimation across different ranges of parameters. We experiment with a linear outcome model ($\beta = 1$), a quadratic model ($\beta = 1$), and a non-polynomial model (labeled as $\beta = \sigma$).

Baselines In our evaluation, we gauge the effectiveness of our proposed method by comparing it against commonly employed estimators such as polynomial regression (Poly), difference-in-means (DM) estimators. Since the polynomial regression model needs exact neighborhoods, we provide them with oracle access to

the true interaction graph.[2]

In Figure 3, we illustrate the relative bias of different estimators across a range of parameters. Figure 3a plots the treatment effect estimate against the strength of interference, 3b is for varying the network size n, and 3c plots it against treatment probability p.

As established in Theorem 1, UNITE models produce unbiased estimates for the linear (first row) and quadratic (second row) outcomes. Note that in Figure 3a, when $r = 0$, there is no interference, and hence most estimators are unbiased. However, when interference increases, baselines suffer from severe bias.

As established in Theorem 2, the self-normalized version UNITE-WIS is biased, but its bias reduces as the number of nodes increases. It also shows a lower variance than UNITE as is common for self-normalized IS estimators. Finally, following Theorem 3, we see that UNITE-DR remains unbiased and shows some variance reduction over the vanilla UNITE estimate.

Effect of General Non-Linearity In the third row of Figure 3, we illustrate results from a non-polynomial model, where the outcomes in this case come from the product of linear and a sigmoidal effect. It can be observed that UNITE shows bias, unlike in the linear and quadratic models where Assumption **A6** holds. However, UNITE still shows a lower bias than other estimators. Further, as it follows from Theorem 7, UNITE's bias reduces with increasing p.

4.2   Case Study: Airbnb

We conduct experiments with a model designed from the AirBnB vacation rentals domain Li et al. (2022). We perform simulations with protocol specified in Brennan et al. (2022). Contrary to the previous experiments, the outcomes here do not follow an explicit exposure mapping. Instead, this simulation works uses a type matching model where if the listing and person have the same type, the probability of acceptance is higher. The measured outcome Y_i is 1 iff the customer successfully books the place. The two different options are considered as recommendation algorithms, and treating each customer increases the application probabilities by a factor of α.

Baselines In this experiment, we evaluate our UNITE estimator against the difference-in-means (DM), as well as two oracle estimator which have access to the true graph. One is the oracle Horvitz-Thompson estimator. The other estimator is EXP (from Brennan et al. (2022)), which assumes both: a

---

[1]For the exact statement we refer the readers to Theorem 3.6 from Ross (2011)

[2]Due to incorporating large neighbourhoods, τ_{HT} failed to yield non-meaningful results in any trial.
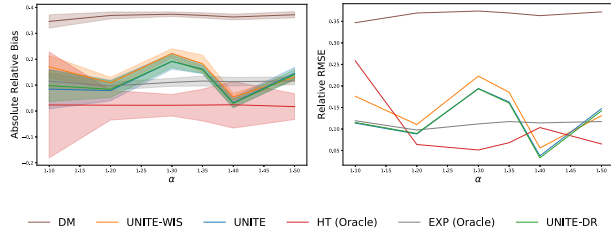
Figure 4: Visualization of relative bias and relative RMSE of different GATE estimators as the indirect treatment effect α increases.

linear exposure model and access to the true graph. The EXP model is equivalent to (Aronow et al., 2017) estimator and computing it requires specifying both the true graph and an exposure function. The relative absolute bias of these estimators is seen in Figure 4.

Since the exposure model can only partly model the actual outcomes, there will be a necessary bias in this case. On the other hand, the Oracle HT estimator gives unbiased though higher variance estimates. From the result it is also clear that our approach works as well as the Oracle Exposure model. Our method has lower bias and similar RMSE compared to the EXP model, while requiring no such information.

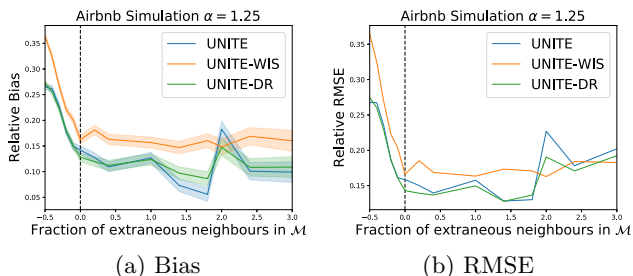### 4.3 Effect of Network Uncertainty

(a) Bias   (b) RMSE

Figure 5: Visualization of the impact of neighbourhood sizes on GATE estimation on the AirBnb Study. Negative fraction of neighbours indicate the case when $\mathcal{M}(i) \subset \mathcal{N}_i$ i.e. we missed pertinent neighbours. The bias tends to be high when given small neighborhoods, as they miss pertinent edges. As the neighborhood sizes increase, the bias reduces, but the uncertainty widens.

Figure 5 illustrates the impact of the neighborhood accuracy $\mathcal{M}(i)$ on estimating $\tau(\vec{0}, \vec{1})$. We experiment with Erdos-Renyi graphs as well as with the AirBnB Model. We fix the interference graph, and compute the treatment effect estimate from our method as we change the assumed neighbourhoods $\mathcal{M}(i)$. In Figure 5, we plot the absolute value of relative bias as

varying proportions of edges are either added or omitted by $\mathcal{M}(i)$ for the AirBnB case. The results from ER graphs are in the Appendix. To maintain simplicity, we maintain uniform $\mathcal{M}(i)$ sizes across all nodes, employing the average number of missed or added edges as the metric along the x-axis.

We see that when $\mathcal{M}(i) \supseteq \mathcal{N}_i$ holds true for all nodes, UNITE is unbiased, and the variance of the estimate increases as the number of extraneous nodes within $\mathcal{M}(i)$ grows. But as expected, when $\mathcal{M}(i)$ misses relevant nodes, the estimate becomes biased, with the overall bias dependent on the influence induced by the missing nodes.

5 CONCLUSION

Network interference exists in many important A/B testing experiments. Our work provides estimators for GATE under a relaxed assumption of having knowledge only about the super-set of neighbors that cause interference. We believe that satisfying this relaxed assumption can be practically far more feasible than requiring the exact network. With both theoretical and experimental analysis, we established the efficacy of our estimator(s) under this assumption.

References

Aittokallio, T. and Schwikowski, B. (2006). Graph-based methods for analysing networks in cell biology. *Briefings in bioinformatics*, 7(3):243–255.

Antman, E. M., Lau, J., Kupelnick, B., Mosteller, F., and Chalmers, T. C. (1992). A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: treatments for myocardial infarction. *Jama*, 268(2):240–248.

Aronow, P. M., Samii, C., et al. (2017). Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912–1947.

Auerbach, E. and Tabord-Meehan, M. (2021). The local approach to causal inference under network interference. *arXiv preprint arXiv:2105.03810*.

Bargagli-Stoffi, F. J., Tortù, C., and Forastiere, L. (2020). Heterogeneous treatment and spillover effects under clustered network interference. *arXiv preprint arXiv:2008.00707*.

Basse, G. W. and Airoldi, E. M. (2018). Model-assisted design of experiments in the presence of network-correlated outcomes. *Biometrika*, 105(4):849–858.

Bhattacharya, R., Malinsky, D., and Shpitser, I. (2020). Causal inference under interference and network uncertainty. In Adams, R. P. and Gogate, V.,

editors, *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pages 1028–1038. PMLR.

Brennan, J., Mirrokni, V., and Pouget-Abadie, J. (2022). Cluster randomized designs for one-sided bipartite experiments. *Advances in Neural Information Processing Systems*, 35:37962–37974.

Cai, J., De Janvry, A., and Sadoulet, E. (2015). Social networks and the decision to insure. *American Economic Journal: Applied Economics*, 7(2):81–108.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.

Chin, A. (2019). Regression adjustments for estimating the global treatment effect in experiments with interference. *Journal of Causal Inference*, 7(2).

Cortez, M., Eichhorn, M., and Yu, C. L. (2022). Graph agnostic estimators with staggered rollout designs under network interference. *Advances in Neural Information Processing Systems*.

Cox, D. R. (1958). Planning of experiments.

D'Amour, A. (2019). On multi-cause causal inference with unobserved confounding: Counterexamples, impossibility, and alternatives. *arXiv preprint arXiv:1902.10286*.

Dorie, V., Harada, M., Carnegie, N. B., and Hill, J. (2016). A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. *Statistics in Medicine*, 35(20):3453–3470.

Eckles, D., Karrer, B., and Ugander, J. (2017). Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference*, 5(1).

Erdos, P., Renyi, A., et al. (1960). On the evolution of random graphs. *Publ. math. inst. hung. acad. sci*, 5(1):17–60.

Fine, P. E. (1993). Herd immunity: history, theory, practice. *Epidemiologic reviews*, 15(2):265–302.

Gui, H., Xu, Y., Bhasin, A., and Han, J. (2015). Network a/b testing: From sampling to estimation. In *Proceedings of the 24th International Conference on World Wide Web*, pages 399–409. International World Wide Web Conferences Steering Committee.

Halloran, M. E. and Hudgens, M. G. (2016). Dependent happenings: a recent methodological review. *Current epidemiology reports*, 3(4):297–305.

Holland, P. W. and Leinhardt, S. (1974). The statistical analysis of local structure in social networks.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685.

Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842.

Kephart, J. O. and White, S. R. (1992). Directed-graph epidemiological models of computer viruses. In *Computation: the micro and the macro view*, pages 71–102. World Scientific.

Kusner, M. J., Sun, Y., Sridharan, K., and Weinberger, K. Q. (2016). Private causal inference. In *Artificial Intelligence and Statistics*, pages 1308–1317. PMLR.

LeSage, J. and Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Leung, M. P. (2016). Treatment and spillover effects under network interference. *Review of Economics and Statistics*, pages 1–42.

Leung, M. P. (2020). Treatment and spillover effects under network interference. *Review of Economics and Statistics*, 102(2):368–380.

Li, H., Zhao, G., Johari, R., and Weintraub, G. Y. (2022). Interference, bias, and variance in two-sided marketplace experimentation: Guidance for platforms. In *Proceedings of the ACM Web Conference 2022*, pages 182–192.

Li, W., Sussman, D. L., and Kolaczyk, E. D. (2021). Causal inference under network interference with noise. *arXiv preprint arXiv:2105.04518*.

Liu, W., Kuramoto, S. J., and Stuart, E. A. (2013). An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research. *Prevention Science*, 14(6):570–580.

Miao, W., Hu, W., Ogburn, E. L., and Zhou, X. (2020). Identifying effects of multiple treatments in the presence of unmeasured confounding. *arXiv preprint arXiv:2011.04504*.

Neyman, J. (1923). On the Application of Probability Theory to Agricultural Experiments: Essay on Principles. *Statistical Science*, 5:465–80. Section 9 (translated in 1990).

Ogburn, E. L., Sofrygin, O., Diaz, I., and Van der Laan, M. J. (2017). Causal inference for social network data. *arXiv preprint arXiv:1705.08527*.

Papadogeorgou, G., Imai, K., Lyall, J., and Li, F. (2020). Causal inference with spatio-temporal data: Estimating the effects of airstrikes on insurgent violence in iraq. *arXiv preprint arXiv:2003.13555*.

Parker, B. M., Gilmour, S. G., and Schormans, J. (2017). Optimal design of experiments on connected units with application to social networks. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(3):455–480.

Pearl, J. (2009). *Causality*. Cambridge university press.

Pouget-Abadie, J., Aydin, K., Schudy, W., Brodersen, K., and Mirrokni, V. (2019). Variance reduction in bipartite experiments through correlation clustering. *Advances in Neural Information Processing Systems*, 32.

Pouget-Abadie, J., Saveski, M., Saint-Jacques, G., Duan, W., Xu, Y., Ghosh, S., and Airoldi, E. M. (2017). Testing for arbitrary interference on experimentation platforms. *arXiv preprint arXiv:1704.01190*.

Randolph, H. E. and Barreiro, L. B. (2020). Herd immunity: understanding covid-19. *Immunity*, 52(5):737–741.

Ranganath, R. and Perotte, A. (2018). Multiple causal inference with latent confounding. *arXiv preprint arXiv:1805.08273*.

Richardson, A., Hudgens, M. G., Gilbert, P. B., and Fine, J. P. (2014). Nonparametric bounds and sensitivity analysis of treatment effects. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, 29(4):596.

Ross, N. (2011). Fundamentals of stein's method. *Probability Surveys*, 8:210–293.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688.

Rubin, D. B. (1987). The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The sir algorithm. *Journal of the American Statistical Association*, 82(398):543–546.

Seshadhri, C., Kolda, T. G., and Pinar, A. (2012). Community structure and scale-free collections of erdos-renyi graphs. *Physical Review E*, 85(5):056109.

Shankar, S., Sinha, R., Mitra, S., Sinha, M., and Fiterau, M. (2023a). Direct inference of effect of treatment (diet) for a cookieless world. In *Proceedings of the The 26th International Conference on Artificial Intelligence and Statistics (AISTATS)*.

Shankar, S., Sinha, R., Mitra, S., Swaminathan, V. V., Mahadevan, S., and Sinha, M. (2023b). Privacy

aware experiments without cookies. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, WSDM '23. Association for Computing Machinery.

Siroker, D. and Koomen, P. (2015). *A/B testing: The most powerful way to turn clicks into customers.* John Wiley & Sons.

Spirtes, P. (2010). Introduction to causal inference. *Journal of Machine Learning Research*, 11(5).

Sussman, D. L. and Airoldi, E. M. (2017a). Elements of estimation theory for causal effects in the presence of network interference. *arXiv preprint arXiv:1702.03578*.

Sussman, D. L. and Airoldi, E. M. (2017b). Elements of estimation theory for causal effects in the presence of network interference. *arXiv preprint arXiv:1702.03578*.

Thomas, P. and Brunskill, E. (2016). Data-efficient off-policy policy evaluation for reinforcement learning. In *International Conference on Machine Learning*, pages 2139–2148. PMLR.

Thomas, P. and Brunskill, E. (2017). Importance sampling with unequal support. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Toulis, P. and Kao, E. (2013). Estimation of causal peer influence effects. In *International Conference on Machine Learning*, pages 1489–1497.

Tukey, H. (1956). Conditional monte carlo for normal samples. In *Proc. Symp. on Monte Carlo Methods*, pages 64–79. John Wiley and Sons.

Ugander, J., Karrer, B., Backstrom, L., and Kleinberg, J. (2013). Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 329–337. ACM.

Veitch, V. and Zaveri, A. (2020). Sense and sensitivity analysis: Simple post-hoc analysis of bias due to unobserved confounding. *arXiv preprint arXiv:2003.01747*.

Viviano, D. (2020). Experimental design under network interference. *arXiv preprint arXiv:2003.08421*.

Wang, Y. and Blei, D. M. (2019). The blessings of multiple causes. *Journal of the American Statistical Association*, 114(528):1574–1596.

Wang, Y., Chakrabarti, D., Wang, C., and Faloutsos, C. (2003). Epidemic spreading in real networks: An eigenvalue viewpoint. In *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings.*, pages 25–34. IEEE.

Wasserman, L. (2006). *All of nonparametric statistics*. Springer Science & Business Media.

Wong, J. C. (2020). Computational causal inference. *arXiv preprint arXiv:2007.10979*.

Yadlowsky, S., Namkoong, H., Basu, S., Duchi, J., and Tian, L. (2018). Bounds on the conditional and average treatment effect with unobserved confounding factors. *arXiv preprint arXiv:1808.09521*.

Yu, C. L., Airoldi, E., Borgs, C., and Chayes, J. (2022). Estimating total treatment effect in randomized experiments with unknown network structure. *Proceedings of the National Academy of Sciences*.

Yuan, Y., Altenburger, K., and Kooti, F. (2021). Causal network motifs: identifying heterogeneous spillover effects in a/b tests. In *Proceedings of the Web Conference 2021*, pages 3359–3370.

A Useful Lemmas

Lemma 1. *Suppose that $\{z_i\}_{i=1..n}$ are mutually independent, with $z_i \sim Bernoulli(p)$. Then, for any set of indices $S, S' \subset [n]$, and function f we have*

$$\mathbb{E}\Big[\prod_{i \in S}\Big(\frac{z_i}{p} - \frac{1-z_i}{1-p}\Big)\prod_{j \in S'} f(z_j)\Big] = \begin{cases} (f(1) - f(0))^{|S \cap S'|}\mathbb{E}[f(z)]^{|S' \setminus S|} & if\ S \subseteq S' \\ 0 & otherwise \end{cases}$$

Proof. Fix S, S'. A given index (node) i can either be only in S or only in S' or in both, with only one of the possibilities being true. Correspondingly the product, $\prod_{i \in S}\Big(\frac{z_i}{p} - \frac{1-z_i}{1-p}\Big)\prod_{j \in S'} f(z_j)$ can be factored into three exclusive products:

$$\prod_{i \in S}\Big(\frac{z_i}{p} - \frac{1-z_i}{1-p}\Big)\prod_{j \in S'} f(z_j) = \prod_{i \in S \setminus S'}\Big(\frac{z_i}{p} - \frac{1-z_i}{1-p}\Big)\prod_{k \in S \cap S'} f(z_k)\Big(\frac{z_k}{p} - \frac{1-z_k}{1-p}\Big)\prod_{j \in S' \setminus S} f(z_j)$$

Applying expectations and noting that z_i are mutually independent, we get:

$$\prod_{i \in S \setminus S'} \mathbb{E}[\frac{z_i}{p} - \frac{1-z_i}{1-p}]\prod_{k \in S \cap S'} \mathbb{E}[f(z_k)\Big(\frac{z_k}{p} - \frac{1-z_k}{1-p}\Big)]\prod_{j \in S' \setminus S}\mathbb{E}f(z_j) = \prod_{i \in S \setminus S'} 0 \prod_{k \in S \cap S'} \frac{\mathbb{E}[z_k f(z_k)] - p\mathbb{E}[f(z_k)]}{p(1-p)}\prod_{j \in S' \setminus S}\mathbb{E}[f(z_j)]$$

The RHS can only be non zero if $S \setminus S' = \{\}$ i.e. $S \subseteq S'$.

Since $\mathbb{E}\Big[f(z_k)\Big(\frac{z_k}{p} - \frac{1-z_k}{1-p}\Big)\Big] = p * f(1) * \frac{1}{p} + (1-p) * f(0) * (\frac{-1}{1-p}) = f(1) - f(0)$; thr RHS when it is non zero simplifies to

$$(f(1) - f(0))^{|S \cap S'|}\mathbb{E}[f(z)]^{|S' \setminus S|}$$

$\square$

Corollary 1. *By putting $f(z) = z$ in Lemma 1we get*

$$\mathbb{E}\Big[\prod_{i \in S}\Big(\frac{z_i}{p} - \frac{1-z_i}{1-p}\Big)\prod_{j \in S'} z_j\Big] = \begin{cases} p^{|S' \setminus S|} & if\ S \subseteq S' \\ 0 & otherwise \end{cases}$$

Lemma 2. *Suppose that $\{z_i\}_{i=1..n}$ are mutually independent, with $z_j \sim Bernoulli(p)$. Then, for any subsets S, S', $\mathbb{E}[\prod_{i \in S} f_i(z_i)\prod_{j \in \S'} \frac{z_j - p}{p}] = \prod_{i \in S \setminus S'} \mathbb{E}[f_i(z_i)]\prod_{k \in S \cap S'}((1-p)(f_k(1) - f_k(0)))\,\mathbb{I}[S' \subseteq S]$*

Proof. Fix S, S'. A given index (node) i can either be only in S or only in S' or in both, with only one of the possibilities being true. Correspondingly the product, $prod_{i \in S} f_i(z_i)\prod_{j \in \S'} \frac{z_j - p}{p}$ can be factored into three exclusive products:

$$\mathbb{E}[\prod_{i \in S} f_i(z_i)\prod_{j \in \S'}\frac{z_j - p}{p}] = \mathbb{E}[\prod_{i \in S \setminus S'} f_i(z_i)\prod_{k \in S \cap S'} f_k(z_k)\frac{z_k - p}{p}\prod_{j \in S' \setminus S}\frac{z_j - p}{p}]$$

$$= \prod_{i \in S \setminus S'} \mathbb{E}[f_i(z_i)]\prod_{k \in S \cap S'}\mathbb{E}[f_k(z_k)\frac{z_k - p}{p}]\prod_{j \in S' \setminus S}\mathbb{E}[\frac{z_j - p}{p}]$$

$$= \prod_{i \in S \setminus S'} \mathbb{E}[f_i(z_i)]\prod_{k \in S \cap S'}((1-p)(f_k(1) - f_k(0)))\prod_{j \in S' \setminus S} 0$$

$$= \prod_{i \in S \setminus S'} \mathbb{E}[f_i(z_i)]\prod_{k \in S \cap S'}((1-p)(f_k(1) - f_k(0)))\,\mathbb{I}[S' \subseteq S]$$

Similarly,

$$\mathbb{E}[\prod_{i \in S} z_i\prod_{j in \S'}\frac{p - z_j}{1-p}] = \prod_{i \in S \setminus S'} \mathbb{E}[f(z_i)]\prod_{k \in S \cap S'}((f(0) - f(1))p)\prod_{j \in S' \setminus S} 0 = \prod_{i \in S \setminus S'} \mathbb{E}[f(z_i)]\prod_{k \in S \cap S'}((f(0) - f(1))p)\,\mathbb{I}[S' \subseteq S]$$

$\square$

Lemma 3. *For any sets $S', \mathcal{M}_i$ such that $S' \subseteq \mathcal{N}_i \subseteq \mathcal{M}_i$*

$$\mathbb{E}\left[\prod_{k \in S'} (z_k) \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1 - p}\right)\right] = 1$$

Proof.

$$\mathbb{E}\left[\prod_{k \in S'} (z_k) \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1 - p}\right)\right] = \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \mathbb{E}\left[\left(\prod_{k \in S'} (z_k) \prod_{j \in S} \frac{z_j - p}{p} - \prod_{k \in S'} (z_k) \prod_{j \in S} \frac{p - z_j}{1 - p}\right)\right]$$

Applying Lemma 2 with $f_i(z) = z_i$ we get

$$= \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left[p^{|S'/S|}(1-p)^{|S' \cap S|}\mathbb{I}[S \subseteq S'] - p^{|S'/S|}(-p)^{|S' \cap S|}\mathbb{I}[S \subseteq S']\right]$$

$$\overset{(b)}{=} \sum_{\substack{S \subseteq S' \\ |S| \leq \beta}} \left[p^{|S'/S|}(1-p)^{|S' \cap S|} - p^{|S'/S|}(-p)^{|S' \cap S|}\right] \tag{S1}$$

(b) follows from that fact that $M_i \supseteq N_i$ for any node i and $\mathbb{I}[S \subseteq S']$ will filter any non subset of S'

$$= \sum_{\substack{S \subseteq S' \\ |S| \leq \beta}} p^{|S'|} \left[p^{-|S|}(1-p)^{|S|} - p^{-|S|}(-p)^{|S|}\right]$$

$$= \sum_{\substack{S \subseteq S' \\ |S| \leq \beta}} p^{|S'|} \left[(\frac{1}{p} - 1)^{|S|} - (-1)^{|S|}\right] \tag{S2}$$

Note that the terms in the sum S2 only depend on sizes of the subset S and not the elements in it. Hence the sum S1 can be rewritten as:

$$= p^{|S'|} \sum_{\substack{r \leq \beta \\ r \leq |S'|}} \binom{|S'|}{r} \left[(\frac{1}{p} - 1)^r - (-1)^r\right]$$

$$= p^{|S'|} \left[\sum_{\substack{r \leq \beta \\ r \leq |S'|}} \binom{|S'|}{r}(\frac{1}{p} - 1)^r - \sum_{\substack{r \leq \beta \\ r \leq |S'|}} \binom{|S'|}{r}(-1)^r\right] \tag{S3}$$

If $|S'| \leq \beta$, we are summing over all subsets of S', and the constraint of $r \leq \beta$ is redundant. Then by applying binomial theorem we get.

$$\overset{(c)}{=} p^{|S'|} \left[\left(1 + (\frac{1}{p} - 1)\right)^{|S'|} - (1 + (-1))^{|S'|}\right] = p^{|S'|}(\frac{1}{p})^{|S'|} = 1$$

where in (c) we used binomial as $\sum_r \binom{n}{r}x^r = (1+x)^n$ □

B   Proof of general β Estimator

Theorem 1. *Under assumptions **A2-5** and $\mathcal{M}_i \supseteq \mathcal{N}_i$, then $\hat{\tau}^\beta$ is unbiased*

Proof. If $Y_i(\boldsymbol{z}) = \sum_{S' \subset \mathcal{N}_i} c_{i,S'} \prod_{j \in S'} \mathbb{I}[z_j = 1]$ then for $\hat{\tau}^\beta$ we get

$$\mathbb{E}[\hat{\tau}^\beta] = \mathbb{E}\left[\frac{1}{n}\sum_i Y_i \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1-p}\right)\right]$$

$$= \mathbb{E}\left[\frac{1}{n}\sum_i \sum_{S' \subset \mathcal{N}_i} c_{i,S'} \prod_{j \in S'} \mathbb{I}[z_j = 1] \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1-p}\right)\right]$$

$$= \frac{1}{n}\sum_i \mathbb{E}\left[\sum_{S' \subset \mathcal{N}_i} c_{i,S'} \prod_{j \in S'} \mathbb{I}[z_j = 1] \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1-p}\right)\right]$$

$$= \frac{1}{n}\sum_i \mathbb{E}\left[\sum_{S' \subset \mathcal{N}_i} c_{i,S'} \prod_{j \in S'} \mathbb{I}[z_j = 1] \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1-p}\right)\right]$$

$$= \frac{1}{n}\sum_i \mathbb{E}\left[\sum_{S' \subset \mathcal{N}_i} c_{i,S'} \prod_{j \in S'} z_j \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1-p}\right)\right]$$

$$= \frac{1}{n}\sum_i \sum_{S' \subset \mathcal{N}_i} c_{i,S'} \underbrace{\mathbb{E}\left[\prod_{j \in S'} z_j \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1-p}\right)\right]}_{\text{E1}}$$

Now applying Lemma 3 on E1 we get

$$= \frac{1}{n}\sum_i \sum_{S' \subset \mathcal{N}_i} c_{i,S'}\,[1] = \tau(\vec{1}, \vec{0})$$

$\square$

Theorem 2. *Estimator $\hat{\tau}_{Lin}$ is the same as $\hat{\tau}^\beta$ for $\beta = 1$*

Proof.

$$\hat{\tau}^\beta|_{\beta=1} = \frac{1}{n}\sum_i Y_i \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq 1}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1-p}\right)$$

$$\frac{1}{n}\sum_i Y_i \sum_{\substack{S = \{k\} \\ k \in \mathcal{M}_i}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1-p}\right)$$

$$= \frac{1}{n}\sum_i Y_i \sum_{j \in \mathcal{M}_i} \left(\frac{z_j - p}{p} - \frac{p - z_j}{1-p}\right)$$

$$= \frac{1}{n}\sum_i Y_i \sum_{j \in \mathcal{M}_i} \left(\frac{z_j}{p} - \cancel{(1)} - \cancel{(-1)} - \frac{1 - z_j}{1-p}\right)$$

$$= \hat{\tau}_{\text{Lin}}$$

$\square$

Together these theorems prove the unbiased part of Theorem 4 and Theorem 1

Theorem 3. *Under assumptions $\textbf{A}2$-5, if $\mathcal{M}_i \not\supseteq \mathcal{N}_i$, then $\hat{\tau}^\beta$ can be biased.*

Proof. We go straight to sum S1 in the proof of Lemma 3. In doing the substituiton in the limits of the summation from $\mathcal{M}_i$ to S' we used the fact that all subsets of S' will be subsets of $\mathcal{M}_i$, and any non-subset of S' is ignored due to $\mathbb{I}[S \subseteq S']$. If $\mathcal{M}_i \not\supseteq \mathcal{N}_i$, then $\exists i^* s.t. \ i^* \in \mathcal{N}_i$ but $i^* \not\subseteq \mathcal{M}_i$. Now take a $S' \subset \mathcal{N}_i$ such that $i* \in S'$. In the summation S1, now the limit of summation goes to subsets of $\mathcal{M}_i$

$$
\mathbb{E}\left[\prod_{k \in S'} (z_k) \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1 - p} \right) \right] = \sum_{\substack{S \subseteq S' \\ S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left[p^{|S'/S|}(1-p)^{|S' \cap S|} - p^{|S'/S|}(-p)^{|S' \cap S|} \right]
$$

$$
= \sum_{\substack{S \subseteq S' \cap \mathcal{M}_i \\ |S| \leq \beta}} \left[p^{|S'/S|}(1-p)^{|S' \cap S|} - p^{|S'/S|}(-p)^{|S' \cap S|} \right]
$$

Following the exact same steps we get:

$$
= p^{|S'|} \sum_{\substack{r \leq \beta \\ r \leq |S'|}} \binom{|S' \cap \mathcal{M}_i|}{r} \left[(\frac{1}{p} - 1)^r - (-1)^r \right]
$$

$$
= p^{|S'|}(\frac{1}{p})^{|S' \cap \mathcal{M}_i|} = p^{|S' \setminus \mathcal{M}_i|} \leq 1 \tag{\textbf{B1}}
$$

Now using definition of $\hat{\tau}^\beta$ we get

$$
\mathbb{E}[\hat{\tau}^\beta] = \mathbb{E}\left[\frac{1}{n} \sum_i Y_i \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1 - p} \right) \right]
$$

$$
= \mathbb{E}\left[\frac{1}{n} \sum_i \sum_{S' \subset \mathcal{N}_i} c_{i,S'} \prod_{j \in S'} \mathbb{I}[z_j = 1] \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1 - p} \right) \right]
$$

$$
= \frac{1}{n} \sum_i \sum_{S' \subset \mathcal{N}_i} c_{i,S'} \mathbb{E}\left[\prod_{j \in S'} \mathbb{I}[z_j = 1] \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1 - p} \right) \right]
$$

which by **B1** is

$$
= \frac{1}{n} \sum_i \sum_{S' \subset \mathcal{N}_i} c_{i,S'} p^{|S' \setminus M_i|} \neq \tau(\vec{1}, \vec{0})
$$

For any S' with $|S'| < \beta$ which influences Y_i but is not a subset of $\mathcal{M}_i$, the corresponding coefficient $c_{i,S'}$ is attenuated by a factor of $p^{|S' \setminus \mathcal{M}_i|}$ in the estimate $\hat{\tau}^\beta$

$\square$

B.1   Variance Analysis

Next, we provide an upper bound for the variance of this estimate. But before we do that we need another helpful Lemma.

Lemma 4. *Suppose that $\{z_i\}_{i=1..n}$ are mutually independent, with $z_j \sim Bernoulli(p)$. Then, for any subsets T, S, S', we have*

$$\mathbb{E}[\prod_{i\in T} f_i(z_i) \prod_{j\in S} \frac{z_j - p}{p} \prod_{j\in S'} \frac{-z_j + p}{1-p}] = \prod_{T\setminus(S\cup S')} \mathbb{E}[f_i] \prod_{(T\cap S)\setminus(S')} (1-p)(f_j(1) - f_j(0)) \prod_{(T\cap S')\setminus(S)} p(f_j(0) - f_j(1)) \quad \times$$

$$\prod_{(T\cap S'\cap S)} -\mathbb{E}[f_i(1 - z_i)] \prod_{(S'\cap S)\setminus T} -1\mathbb{I}[S\Delta S' \subseteq T]$$

Proof. We first apply Lemma 2 on the above expression, and see that it is non-zero only if $S \subseteq (T \cup S')$. We can also apply the second result from Lemma 2 and get that the expression is non-zero only if $S' \subseteq (T \cup S)$. Since $S \cap S' \subseteq S, S'$, the two conditions imply that for the expression to be non-zero we need $S\Delta S' \subseteq T$

Next we need to identify the expected value of the combined function applied on a node as in Lemma 2.

- For indices in $T \setminus (S \cup S')$ we get $\mathbb{E}[f_i]$

- For indices in $S \setminus (T \cup S')$ we get 0. This is subsumed in $S\Delta S' \subseteq T$

- For indices in $S' \setminus (T \cup S)$ we get 0. This is subsumed in $S\Delta S' \subseteq T$

- For indices in $(S \cap T) \setminus S'$ we get $\mathbb{E}[f_i \frac{z_j - p}{p}] = (1-p)(f_i(1) - f_i(0))$

- For indices in $(S' \cap T) \setminus S$ we get $\mathbb{E}[f_i \frac{-z_j + p}{1-p}] = p(f_i(0) - f_i(1))$

- For indices in $(S' \cap S) \setminus T$ we get $\mathbb{E}[\frac{z_j - p}{p} \frac{-z_j + p}{1-p}] = -1$

- For indices in $T \cap S \cap S'$ we get $\mathbb{E}[f_i \frac{z_j - p}{p} \frac{-z_j + p}{1-p}] = -pf_i(0) - (1-p)f_i(1) = -\mathbb{E}[f_i(1 - z_i)]$

$\square$

Theorem 4. *Under assumptions **A**2-5 and assuming $\mathcal{M}_i \supseteq \mathcal{N}_i$, we have that $\mathrm{Var}[\hat{\tau}^\beta] \leq \frac{1}{n} Y_{\max}^2 (p(1-p))^{-\beta} d_{\mathcal{M}}^{2\beta} d_{\mathcal{N}}^2$*

Proof. For simplicity of notation we denote by $\Psi(S)$ and $\Psi'(S)$ the products $\prod_{i\in S} \frac{z_i - p}{p}$ and $\prod_{i\in S} \frac{p - z_i}{1-p}$. We would also first ignore the additional noise ε while presenting the key derivation.

For any random variable X, $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$ Applying the same on $\hat{\tau}^\beta$ we get,

$$\mathrm{Var}[\hat{\tau}^\beta] \leq \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n Y_i \sum_{\substack{S\subseteq\mathcal{M}_i \\ |S|\leq\beta}} \left(\prod_{j\in S} \frac{z_j - p}{p} - \prod_{j\in S} \frac{p - z_j}{1-p}\right)\right)^2\right]$$

$$= \frac{1}{n^2}\mathbb{E}\left[\sum_{i,i'=1}^n Y_i Y_{i'} \sum_{\substack{S\subseteq\mathcal{M}_i \\ |S|\leq\beta}} \left(\prod_{j\in S} \frac{z_j - p}{p} - \prod_{j\in S} \frac{p - z_j}{1-p}\right) \sum_{\substack{S'\subseteq\mathcal{M}_{i'} \\ |S'|\leq\beta}} \left(\prod_{j'\in S'} \frac{z_{j'} - p}{p} - \prod_{j'\in S} \frac{p - z_{j'}}{1-p}\right)\right]$$

$$= \frac{1}{n^2}\mathbb{E}\left[\sum_{i,i'=1}^n Y_i Y_{i'} \sum_{\substack{S\subseteq\mathcal{M}_i \\ |S|\leq\beta}} \left(\Psi(S) - \Psi'(S)\right) \sum_{\substack{S'\subseteq\mathcal{M}_{i'} \\ |S'|\leq\beta}} \left(\Psi(S') - \Psi'(S')\right)\right]$$

Putting in the value of Y_i, we get

$$
= \frac{1}{n^2} \mathbb{E} \left[\sum_{\substack{i,i'=1}}^{n} \sum_{\substack{T \subseteq \mathcal{N}_i \\ |T| \leq \beta}} c_{i,T} Z(T) \sum_{\substack{T' \subseteq \mathcal{N}_{i'} \\ |T'| \leq \beta}} c_{i,T'} Z(T') \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \Big(\Psi(S) - \Psi'(S) \Big) \sum_{\substack{S' \subseteq \mathcal{M}_{i'} \\ |S'| \leq \beta}} \Big(\Psi(S') - \Psi'(S') \Big) \right]
$$

$$
= \frac{1}{n^2} \sum_{\substack{i,i'=1}}^{n} \sum_{\substack{T \subseteq \mathcal{N}_i \\ |T| \leq \beta}} c_{i,T} \sum_{\substack{T' \subseteq \mathcal{N}_{i'} \\ |T'| \leq \beta}} c_{i,T'} \mathbb{E} \left[Z(T) Z(T') \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \Big(\Psi(S) - \Psi'(S) \Big) \sum_{\substack{S' \subseteq \mathcal{M}_{i'} \\ |S'| \leq \beta}} \Big(\Psi(S') - \Psi'(S') \Big) \right]
$$

$$
= \frac{1}{n^2} \sum_{\substack{i,i'=1}}^{n} \sum_{\substack{T \subseteq \mathcal{N}_i \\ |T| \leq \beta}} c_{i,T} \sum_{\substack{T' \subseteq \mathcal{N}_{i'} \\ |T'| \leq \beta}} c_{i,T'} \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \sum_{\substack{S' \subseteq \mathcal{M}_{i'} \\ |S'| \leq \beta}} \mathbb{E} \left[Z(T) Z(T') \Big(\Psi(S) - \Psi'(S) \Big) \Big(\Psi(S') - \Psi'(S') \Big) \right]
$$

$$
= \frac{1}{n^2} \sum_{\substack{i,i'=1}}^{n} \sum_{\substack{T \subseteq \mathcal{N}_i \\ |T| \leq \beta}} c_{i,T} \sum_{\substack{T' \subseteq \mathcal{N}_{i'} \\ |T'| \leq \beta}} c_{i,T'} \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \sum_{\substack{S' \subseteq \mathcal{M}_{i'} \\ |S'| \leq \beta}} \mathbb{E} \Big[Z(T) Z(T') \Big(\Psi(S) \Psi(S') + \Psi'(S) \Psi'(S') +
$$

$$
- \Psi(S) \Psi'(S') - \Psi'(S) \Psi'(S) \Big) \Big]
$$

By applying Lemma , we consider in the expectation only the terms with sets S, S', T, T' such that $S \Delta S' \subseteq T \cup T'$. Moreover when these values are non-zero, they are upper bounded by $(p(1-p))^{-\beta}$

Additionally we have the following

$$
\mathbb{E}[Z(T) Z(T') \Psi(S) \Psi(S')] = p^{|T'' \setminus (S \cup S')|} \big(\frac{(1-p)^2}{p} \big)^{|T'' \cap S \cap S'|} \big(\frac{1-p}{p} \big)^{|(S' \cap S) \setminus T''|} \mathbb{I}[S \Delta S' \subseteq T'']
$$

$$
\leq \big(\frac{1-p}{p} \big)^{|(S' \cap S) \setminus T''|} \mathbb{I}[S \Delta S' \subseteq T'']
$$

Similarly,

$$
\mathbb{E}[Z(T) Z(T') \Psi'(S) \Psi'(S')] \leq \big(\frac{(p)}{1-p} \big)^{|(S' \cap S) \setminus T''|} \mathbb{I}[S \Delta S' \subseteq T'']
$$

$$
\mathbb{E}[Z(T) Z(T') \Psi(S) \Psi'(S')] = p^{|T'' \setminus (S \cup S')|} (1-p)^{|T'' \cap S \cap S'|} (-1)^{|(S' \cap S) \setminus T''|} \mathbb{I}[S \Delta S' \subseteq T''] \in (-1, 1)
$$

where $T'' = T \cup T'$

Since the first two terms $\mathbb{E}[Z(T) Z(T') \Psi(S) \Psi(S')]$ $\mathbb{E}[Z(T) Z(T') \Psi'(S) \Psi'(S')]$ add together in the variance, this leads to a bound of $\Big((\frac{1-p}{p})^{|(S' \cap S) \setminus T''|} + (\frac{(p)}{1-p})^{|(S' \cap S) \setminus T''|} \Big) \mathbb{I}[S \Delta S' \subseteq T'']$.

Next note that none of the sets T, T', S, S' can be bigger than β elements. Under this constraint the bound in the previous equation is maximized by setting $S = S'$ and $S \cap S'$ disjoint from T''. Putting the corresponding sizes of the sets, we get a bound of

$$
\Big((\frac{1-p}{p})^{\beta} + (\frac{(p)}{1-p})^{|\beta|} \Big) \leq (p(1-p))^{-\beta}
$$

Putting these in the equation for variance we get

$$
\mathrm{Var}[\hat{\tau}^{\beta}] \leq \frac{1}{n^2} \sum_{i,i'} \sum_{\substack{T \subseteq \mathcal{N}_i \\ |T| \leq \beta}} c_{i,T} \sum_{\substack{T' \subseteq \mathcal{N}_{i'} \\ |T'| \leq \beta}} c_{i,T'} \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \sum_{\substack{S' \subseteq \mathcal{M}_{i'} \\ |S'| \leq \beta}} (p(1-p))^{-\beta} \mathbb{I}[S \Delta S' \subseteq T \cup T']
$$

Let $\mathcal{M}_i^2$ denote the set of indices j that are neighbours of neighbours of i as per model $\mathcal{M}$. Similarly, one can consider $\mathcal{N}_i^2$ and $d_{\mathcal{N}}$. If $i' \notin \mathcal{M}_i^2$ then the sets S, S' in the expectation have no overlap, and hence can be ignored. Furthermore $|\mathcal{M}_i^2| \leq d_{\mathcal{M}}^2$, where $d_{\mathcal{M}}$ is the max degree of a node. Similarly, the coefficients c can only occur from $\mathcal{N}^2$. We also see that the number of T, T' is bounded by $d_{\mathcal{N}}^{\beta}$. Finally for any cluter of sets S, S', T, T' that contribute, they can be enumerated by choosing a set corresponding to $T \cup T'$ of size k, and seperately counting the elements in S, S'. Hence this sum is upper bounded by

$$\text{Var}[\hat{\tau}^{\beta}] \leq \frac{1}{n^2} \sum_i Y_{\max}^2 \sum_{i' \in \mathcal{N}_i^2} \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \sum_{\substack{S' \subseteq \mathcal{M}_{i'} \\ |S'| \leq \beta}} (p(1-p))^{-\beta} \leq \frac{1}{n} Y_{\max}^2 (p(1-p))^{-\beta} d_{\mathcal{M}}^{\beta+1} d_{\mathcal{N}}^{\beta+1} \tag{7}$$

Next, we add back the additional uncertainty in Y_i due to ε. If ε is an entirely independent error with variance σ^2, we get an additional term of the form $\mathbb{E}\left[\varepsilon^2\left(\Psi(S)\Psi(S') + \Psi'(S)\Psi'(S') - \Psi(S)\Psi'(S') - \Psi'(S)\Psi'(S)\right)\right]$. By replace $c_{i,T/T'}$ by ε, and working through the same derivation we have this additional term being bounded above by $\frac{\sigma^2}{n} d_{\mathcal{M}}^{\beta+1} d_{\mathcal{N}}^{\beta+1} (p(1-p))^{-\beta}$

If $Y_{\max}$ and $d_{\mathcal{M}}$ are bounded, then the above variance asymptotes to 0 as $n \to \infty$. Thus as $n \to \infty$, the estimator $\hat{\tau}^{\beta}$ is a.s. constant. Thus the previous theorem on variance bound combined with the earlier proves unbiasedness shows the asymptotic consistency of the estimators $\hat{\tau}_{Lin}$ and $\hat{\tau}^{\beta}$. (i.e. Theorems 4 and 1)

$\square$

Theorem 5. *(same as Theorem 7 of main text) For the non-linear model 6, if g_i is $k+1$ times differentiable and the $k+1$th derivative is bounded by C, then the absolute bias*

$$\left|\mathbb{E}[\hat{\tau}^{\beta}] - \tau(\vec{1}, \vec{0})\right| \leq \frac{C \max(p, 1-p)}{(k+1)!}.$$

Proof. By Taylor's theorem $Y_i(\boldsymbol{z})$ can be written as a polynomial of order k in $\boldsymbol{z}[\mathcal{N}_i]$ with a $k+1$ order residual term

$$Y_i(\boldsymbol{z}) = g_i^0 + \sum_{z_j} \partial_{z^j} g_i z_j + \sum_{z_j, z_k} \partial_{z_j, z_k}^2 g_i z_j z_k + .. + R_{k+1} = Y_{i,k}(\boldsymbol{z}) + R_{k+1}$$

where $R_{k+1} = \frac{\partial^{k+1}}{k+1!} Y_i(z^*) poly(\boldsymbol{z}[\mathcal{N}_i], k+1)$ for some $z^* \in [0,1]^{|\mathcal{N}_i|}$

For simplicity we denote $\left(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1-p}\right)$ as Ψ_S) Now we use the result S2 from the previous theorem with S' being a set of size $k+1$ and $\beta = k$ we get

$$\mathbb{E}[z_1 z_2 .. z_{k+1} \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq k}} \Psi_S] = p^{k+1} \sum_{r \leq k} \binom{k+1}{r} \left[(\frac{1}{p} - 1)^r - (-1)^r\right]$$

$$= p^{k+1}\left(\frac{1}{p}^{k+1} - (\frac{1}{p} - 1)^{k+1} + (-1)^{k+1}\right)$$

Since $Y_{i,k}(\boldsymbol{z})$ is estimated without bias by our method, the bias of using the k order approximated is given by Hence

$$|\mathbb{E}[R_{k+1}\sum_{\substack{S\subseteq\mathcal{M}_i \\ |S|\leq k}}\Psi_S)]| = |\mathbb{E}[\frac{\partial^{k+1}}{k+1!}Y_i(z^*)poly(\boldsymbol{z}[\mathcal{N}_i],k+1)\Psi_S]|$$

$$\leq |\frac{C}{k+1!}|\mathbb{E}[poly(\boldsymbol{z}[\mathcal{N}_i],k+1)\Psi_S]|$$

$$\leq |\frac{C}{k+1!}|p^{k+1}\left(\frac{1}{p}^{k+1}-(\frac{1}{p}-1)^{k+1}+(-1)^{k+1}\right)|$$

$$= O|\frac{Cp}{k+1!}|$$

Similarly $Y_i(\boldsymbol{z})$ can be expanded around $\vec{1}$ and a similar derivation follows with z_i replaced by $1-z_i$, and p replaced by $1-p$. Putting both together derives the bound in the theorem □

C Proof of Self-Normalized and Doubly Robust Estimators

Since the linear estimator is just a subcase of the general estimate $\hat{\tau}_{WIS}^{\beta}$, we focus on the more general version.

Theorem 6. *Under assumptions **A2-5** and $\mathcal{M}_i \supseteq \mathcal{N}_i$, then $\hat{\tau}_{WIS}^{\beta}$ is asymptotically consistent*

Proof. Let $\rho_i = z_j/p$ and $\rho_i' = \frac{1-z_j}{1-p}$. Moreover let $\bar{\rho}_j$ and $\bar{\rho}_j'$ be the corresponding self normalized values of ρ and ρ', as in 4. The self-normalized estimate for $\hat{\tau}^{\beta}$ is given by:

$$\hat{\tau}_{\text{WIS}}^{\beta} = \frac{1}{n}\sum_{i=1}^{n}Y_i\sum_{\substack{S\subseteq\mathcal{M}_i \\ |S|\leq\beta}}\left(\prod_{j\in S}(\bar{\rho}_j-1)-\prod_{j\in S}(-1+\bar{\rho}_j')\right)$$

Note that as all z_i are independent, so are all ρ,ρ'. As such we can apply Kolmogorove's strong law of large numbers to any linear combination of these variables and get the following

$$\lim_{n\to\infty}\frac{1}{n}\sum_j\sum_{k\in M_j}\frac{1}{\mathcal{M}_k'}\rho_k \xrightarrow{a.s.} \sum_j\sum_{k\in M_j}\frac{1}{\mathcal{M}_k'}\mathbb{E}[\rho_k]=1 \qquad \lim_{n\to\infty}\frac{1}{n}\sum_j\sum_{k\in M_j}\frac{1}{\mathcal{M}_k'}\rho_k' \xrightarrow{a.s.} \sum_j\sum_{k\in M_j}\frac{1}{\mathcal{M}_k'}\mathbb{E}[\rho_k']=1$$

Since $1/t$ is a continuous function at $t = 1$, we can apply Slutsky continuous mapping theorem on 4 to get that

$$\text{as}\lim_{n\to\infty}: \quad \bar{\rho}_i \to \rho_i \text{ and } \bar{\rho}'_i \to \rho_i'.$$

This proves the consistency of $\hat{\tau}_{WIS}^{\beta}$ (Theorems 2 and 5 of main text). Moreover as $\bar{\rho}\neq\rho$, the expected value of $\hat{\tau}_{WIS}^{\beta}$ will not generally equal $\hat{\tau}^{\beta}$. □

Next we demonstrate the unbiasedness and consistency of the DR estimator. Once again we focus on the general β case, as the linear case is a corollary.

Theorem 7. *Under assumptions **A2-5**, and assuming $\mathcal{M}_i \supseteq \mathcal{N}_i$, $\mathbb{E}[\hat{\tau}_{DR}^{\beta}] = \tau(\vec{1},\vec{0})$ and $\hat{\tau}_{DR}^{\beta} \xrightarrow{a.s.} \tau(\vec{1},\vec{0})$.*

Proof. Since $X \perp\!\!\!\perp Z$ $\mathbb{E}[f(X)g(Z)] = \mathbb{E}[f(X)]\mathbb{E}[g(Z)]$. Applying this on Equation 5, we get

$$
\mathbb{E}[\hat{\tau}_{DR}^\beta] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \tilde{Y}_i \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S}(\rho_j - 1) - \prod_{j \in S}(-1 + \rho_j')\right)\right] + \frac{1}{n}\sum_i \mathbb{E}[f_1(X_i) - f_0(X_i)]
$$

$$
= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n (Y_i - z_i(f_1(X_i)) - (1-z_i)f_0(X_i)) \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\prod_{j \in S}(\rho_j - 1) - \prod_{j \in S}(-1 + \rho_j')\right)\right]
$$

$$
+ \frac{1}{n}\sum_i \mathbb{E}[f_1(X_i) - f_0(X_i)]
$$

$$
\overset{a}{=} \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n Y_i \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\Psi(S) - \Psi'(S)\right)\right] - \mathbb{E}[(f_1(X))]\mathbb{E}[z_i \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\Psi(S) - \Psi'(S)\right)]
$$

$$
- \mathbb{E}[(f_0(X))]\mathbb{E}[(1-z_i) \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \left(\Psi(S) - \Psi'(S)\right)] + \frac{1}{n}\sum_i \mathbb{E}[f_1(X_i) - f_0(X_i)]
$$

$$
= \tau(\vec{1}, \vec{0}) - \mathbb{E}[f_0(X)](1) - \mathbb{E}[f_1(X)](-1) + \frac{1}{n}\sum_i \mathbb{E}[f_1(X_i) - f_0(X_i)] = \tau(\vec{1}, \vec{0})
$$

Effectively, $\hat{\tau}_{DR}$, replaces $c_{i,\{i\}}$ by $c_{i,\{i\}} - f_{z_i}(X)$. As long as these $f_{0/1}(X)$ are bounded, the same analysis of variance holds with the modified c. Thus $\hat{\tau}_{DR}^\beta$ inherits consistency from $\hat{\tau}^\beta$ $\qquad\square$

Proposition 1. *If the outcome models $f(X)$ are accurate, the DR estimator $\hat{\tau}_{DR}^\beta$ have lower variance than $\hat{\tau}^\beta$.*

This can be seen from the bound for variance in Eqaution 7. Note that the bound is in terms of the variance of the independent noise σ and the magnitude of the outcome $\max |Y|$. The DR estimates center the outcomes at the outcome models, $f(X)$, the bound for the DR estimator is determined by $\max |Y - f(X)|$, which if the outcome models are good should be smaller than $\max |Y|$.

This term provides support for the idea that with a good model so that $Y_i - f(X_i)$ has smaller coefficients c_i (and hence lower $Y_{\max}$) or a lower error in ε, the variance of the DR estimator can be lower than the standard estimate.

D   Relationship with HT Estimator

$$\hat{\tau}_{\mathrm{HT}} = \frac{1}{n} \sum_i Y_i \left(\prod_{j \in \mathcal{N}_i} \frac{z_j}{p} - \prod_{j \in \mathcal{N}_i} \frac{(1 - z_j)}{(1 - p)} \right) \tag{8}$$

$$\mathbb{E}[\hat{\tau}_{\mathrm{HT}}] = \frac{1}{n} \sum_i \mathbb{E} \left[Y_i \left(\prod_{j \in \mathcal{N}_i} \frac{z_j}{p} - \prod_{j \in \mathcal{N}_i} \frac{(1 - z_j)}{(1 - p)} \right) \right] \tag{9}$$

$$= \frac{1}{n} \sum_i \mathbb{E} \left[\left(\sum_{k \in \mathcal{N}_i} c_{ik} \mathbb{I}[z_k = 1] \right) \left(\prod_{j \in \mathcal{N}_i} \frac{z_j}{p} - \prod_{j \in \mathcal{N}_i} \frac{(1 - z_j)}{(1 - p)} \right) \right] \tag{10}$$

$$= \frac{1}{n} \sum_i \sum_{k \in \mathcal{N}_i} \mathbb{E} \left[c_{ik} \mathbb{I}[z_k = 1] \left(\prod_{j \in \mathcal{N}_i} \frac{z_j}{p} - \prod_{j \in \mathcal{N}_i} \frac{(1 - z_j)}{(1 - p)} \right) \right] \tag{11}$$

$$= \frac{1}{n} \sum_i \sum_{k \in \mathcal{N}_i} \mathbb{E} \left[c_{ik} \mathbb{I}[z_k = 1] \frac{z_k}{p} \right] \underbrace{\mathbb{E} \left[\prod_{j \in \mathcal{N}_i, j \neq k} \frac{z_j}{p} \right]}_{=1} - \frac{1}{n} \sum_i \sum_{k \in \mathcal{N}_i} \mathbb{E} \left[c_{ik} \mathbb{I}[z_k = 1] \frac{1 - z_k}{1 - p} \right] \underbrace{\mathbb{E} \left[\prod_{j \in \mathcal{N}_i, j \neq k} \frac{(1 - z_j)}{(1 - p)} \right]}_{=1} \tag{12}$$

$$= \frac{1}{n} \sum_i \sum_{k \in \mathcal{N}_i} c_{ik} \mathbb{E} \left[\mathbb{I}[z_k = 1] \left(\frac{z_k}{p} - \frac{(1 - z_k)}{(1 - p)} \right) \right] \tag{13}$$

$$= \mathbb{E} \left[\frac{1}{n} \sum_i \sum_{k \in \mathcal{N}_i} c_{ik} \mathbb{I}[z_k = 1] \left(\frac{z_k}{p} - \frac{(1 - z_k)}{(1 - p)} \right) \right] \tag{14}$$

$$\overset{(a)}{=} \mathbb{E} \left[\frac{1}{n} \sum_i \left(\sum_{k \in \mathcal{N}_i} c_{ik} \mathbb{I}[z_k = 1] \right) \left(\sum_{k \in \mathcal{N}_i} \frac{z_k}{p} - \frac{(1 - z_k)}{(1 - p)} \right) \right] \tag{15}$$

$$= \mathbb{E} \left[\underbrace{\frac{1}{n} \sum_i Y_i \left(\sum_{k \in \mathcal{N}_i} \frac{z_k}{p} - \frac{(1 - z_k)}{(1 - p)} \right)}_{\hat{\tau}_{\mathrm{Lin}}} \right] \tag{16}$$

where (a) follows from observing that the sum of cross product terms is 0,

$$\mathbb{E} \left[\sum_{k \in \mathcal{N}_i} c_{ik} \mathbb{I}[z_k = 1] \sum_{j \in \mathcal{N}_i, j \neq k} \frac{z_j}{p} - \frac{(1 - z_j)}{(1 - p)} \right) \right] = \mathbb{E} \left[\sum_{k \in \mathcal{N}_i} c_{ik} \mathbb{I}[z_k = 1] \mathbb{E} \left[\underbrace{\sum_{j \in \mathcal{N}_i, j \neq k} \frac{z_j}{p} - \frac{(1 - z_j)}{(1 - p)}}_{=0} \right] \right] = 0. \tag{17}$$

E   Additional Experiment Results

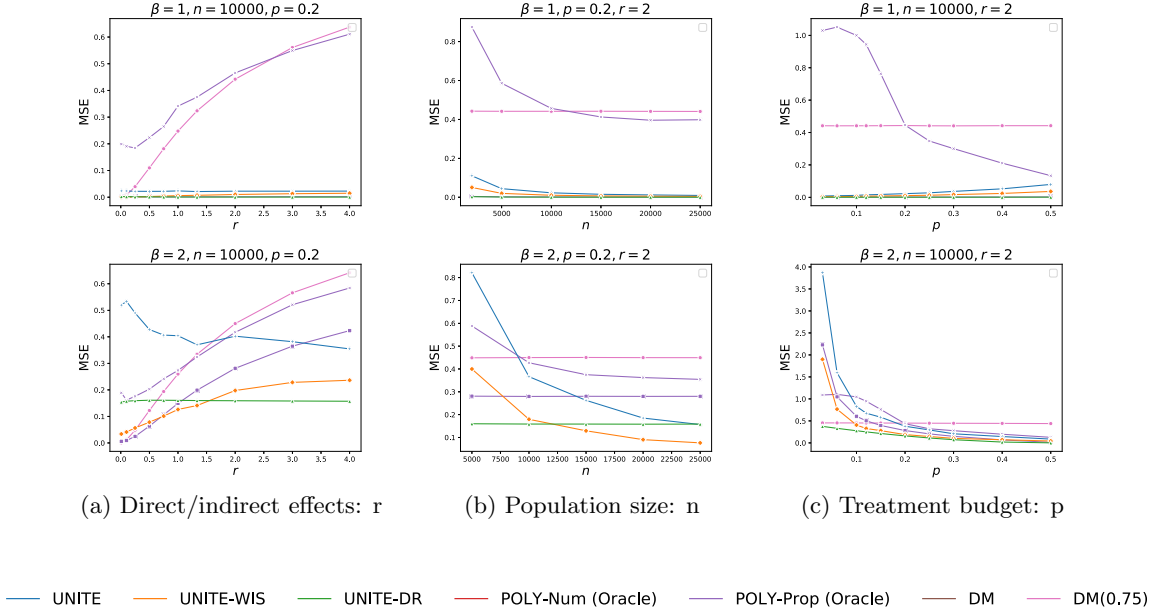(a) Direct/indirect effects: r   (b) Population size: n   (c) Treatment budget: p

Figure 6: Performance of GATE estimators under Bernoulli design on Erdos-Renyi networks. Y-axes represent the MSE. The rows correspond to different generative model for the potential outcomes. Columns correspond to different parameters being varied: (a) Strength of interference, where the x-axis corresponds to the average ratio of indirect to direct effects $r = \frac{1}{n}\sum_{i,j}\frac{|c_{ij}|}{c_{ii}}$. (b) Population size, where the x-axis corresponds to the number of nodes n (c) Treatment budget, where the x-axis corresponds to the probability of treatment 1 (p).
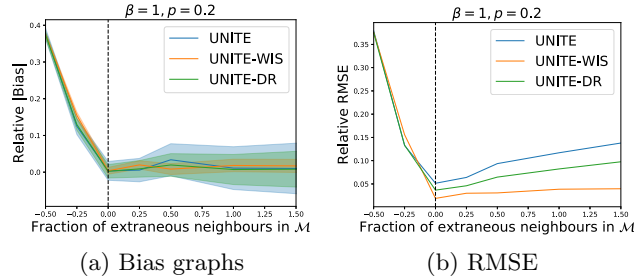

(a) Bias graphs   (b) RMSE

Figure 7: Visualization of the impact of neighbourhood sizes on GATE estimation. Negative fraction of neighbours indicate the case when $\mathcal{M}(i) \subset \mathcal{N}_i$ i.e. we missed pertinent neighbours. The bias tends to be high when gives small neighbourhoods, as they miss pertinent edges. As the neighbourhood sizes increase, the bias reduces, but the uncertainty widens.

F   Checklist

1. For all models and algorithms presented, check if you include:

   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]

   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]

   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

2. For any theoretical claim, check if you include:

(a) Statements of the full set of assumptions of all theoretical results. [Yes]

(b) Complete proofs of all theoretical results. [Yes]

(c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:

   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]

   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]

   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]

   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [No]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

   (a) Citations of the creator If your work uses existing assets. [Yes]

   (b) The license information of the assets, if applicable. [Not Applicable]

   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]

   (d) Information about consent from data providers/curators. [Not Applicable]

   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. [Not Applicable]

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]