Experimentation under Treatment Dependent Network Interference

Shiv Shankar[1] **Ritwik Sinha**[2] **Madalina Fiterau**[1]

[1]University of Massachusetts, Massachusetts, USA
[2]Adobe Research, California, USA,

Abstract

Randomised Controlled Trials (RCTs) are a fundamental aspect of data-driven decision-making. RCTs often assume that the units are not influenced by each other. Traditional approaches addressing such effects assume a fixed network structure between the interfering units. However, real-world networks are rarely static, and treatment assignments can actively reshape the interference structure itself, as seen in financial access interventions that alter informal lending networks or healthcare programs that modify peer influence dynamics. This creates a novel and unexplored problem: estimating treatment effects when the interference network is determined by treatment allocation. In this work, we address this gap by proposing two single-experiment estimators for scenarios where network edges depend on nodal treatments constructed from instrumental variables derived from neighbourhood treatments. We prove their unbiasedness and experimentally validate the proposed estimators both on synthetic and real data.

1 INTRODUCTION

Randomized controlled trials (RCTs), or A/B testing, is a fundamental tool for assessing the effectiveness of interventions across multiple disciplines, including healthcare[Antman et al., 1992], and digital platforms[Siroker and Koomen, 2015]. In such a test, treatment (group A) and control (group B) assignments are made independently of other variables, including potentially unknown ones. The outcomes from the two groups can be compared to estimate the desired causal effects. Such an experimentation-based approach empowers data-driven decision-making about the most effective treatments [Aral and Walker, 2011].

Despite its basic soundness, A/B testing is not without challenges, particularly in large-scale experiments where key assumptions may not hold [Pouget-Abadie, 2018, Shankar et al., 2025]. One major issue is interference between subjects, where individuals in the control group are indirectly affected by the treatment assigned to others. This spillover can distort the estimated treatment effect and lead to biased conclusions. For instance, in social networks, recommendations made to users in the treatment group may be shared with those in the control group, reducing the observed difference between the two groups [Brennan et al., 2022, Pouget-Abadie et al., 2017]. Similarly, in public health studies, herd immunity can lead to a spillover effect, making it challenging to isolate the direct impact of a vaccination program [Randolph and Barreiro, 2020, Fine, 1993].

This phenomenon, where treatment of a unit affects outcomes for other units, has been studied in the causal literature [Hudgens and Halloran, 2008, LeSage and Pace, 2009] under the name of interference. A common assumption in such studies is that the structure of interference is encoded by an apriori known network [Ogburn et al., 2017, Leung, 2020]. This is the *neighbourhood interference* assumption, where interference is confined within neighbours in a graph. This dependence graph is typically inferred using observable data like social connections [Aronow et al., 2017], historical user interactions [Bakshy et al., 2012, Karrer et al., 2021]) or from a user linking model [Sinha et al., 2014, Saha Roy et al., 2015].

However, in practice, the network structure obtained for post-experimentation analysis is rarely static [Heckman and Pinto, 2015, Sävje, 2024, Sweet and Adhikari, 2020]. Furthermore, the interference graph itself may be affected by the treatment [Gao, 2024, Rogowski and Sinclair, 2012]. A classic example comes from a case study of the introduction of financial and banking access to households in an underdeveloped village [Prina, 2015]. In such communities, families and friends often serve as informal lenders when facing financial hardships. However, the introduction of banking access can lead to changes in these informal lending connections. For example, those units with access

to banks may not borrow from each other as previously. On the other hand, there may be increased lending between peers among whom only one has bank access. Similarly, in healthcare interventions, individuals encouraged to join peer support groups will experience different levels of social influence than those unaware of such networks, making the interference structure dynamic rather than fixed [Arminen, 1998]. These cases *introduce a new scenario, requiring the estimation of treatment effects when the network structure of interference depends on the treatment allocation*.

Contribution . In this work, we consider a network interference scenario in which the existence of edges between nodes is determined by the treatments assigned to those nodes. We provide two different single-experiment estimators for this problem. We show them to be unbiased and experimentally validate their performance.

2 RELATED WORK

Network Interference Network interference is a well-studied topic in causal inference literature [Basse and Airoldi, 2018, Cai et al., 2015, Chin, 2019, Gui et al., 2015, Toulis and Kao, 2013]. First, formally identified by Cox [1958], interference relates to a violation of the Stable Unit Treatment Value Assumption (SUTVA) [Rubin, 1978]. Network interference [Hudgens and Halloran, 2008] relates to the idea that the effects on a unit can be encapsulated in a neighbourhood structure. Most approaches include assumptions about the interference neighbourhood [Bargagli-Stoffi et al., 2020, Frank and Xu, 2020]. Hudgens and Halloran [2008] proposed a method based on clustered interference, which was later extended by Zhang et al. [2023], Ogburn et al. [2024], Shankar et al. [2024a] to allow more flexible network structures. Some other methods focus on using graphical causal models to directly adjust for interference [Ogburn and VanderWeele, 2014, Spohn et al., 2023, Shpitser et al., 2017]. Shankar et al. [2024b] have extended the work on interference to other distributional quantities such as median and CVar. Linear interference model [Sussman and Airoldi, 2017, Jiang and Wang, 2023, Pouget-Abadie, 2018] or exposure mappings [Aronow et al., 2017, Sävje et al., 2021a] are common assumptions for incorporating heterogeneity in interference. O'Riordan and Gilligan-Lee [2025] extend methods based on linearity assumptions to include semi-parametric models. We summarize some common approaches and how our method differs from them in Table 1. A detailed discussion of these is in the Appendix.

Misspecified and Uncertain Interference A major challenge in network interference analysis is dealing with noisy or misspecified networks [Carroll et al., 2006, Ogburn and Vanderweele, 2013, Lockwood and McCaffrey, 2016]. Recently, some methods have been developed to handle the strong assumptions often made in network interference literature (e.g., Leung [2022], Wang et al. [2020], Sävje [2024], Auerbach et al. [2024], Shankar et al. [2023b]). In a related direction, research has also focused on settings where the underlying network structure is unknown or only partially known (e.g., Chin [2019], Sävje et al. [2021a], Cortez-Rodriguez et al. [2023], Shankar et al. [2025]). Most of these methods are based on multiple measurements [Shankar et al., 2023b, Cortez et al., 2022, Yu et al., 2022], though some other approaches exist based on outcome assumptions [Shankar et al., 2024c] and on uncertainty estimates for the network structure [Zhang et al., 2023]. Other approaches include methods based on measurement error [Miao et al., 2018, Kuroki and Pearl, 2014] and confounding models [Shpitser et al., 2021]. When networks are uncertain, methods for obtaining partial identification bounds for treatment effects have been proposed [Zhao et al., 2017, Yadlowsky et al., 2018].

However, these methods still assume a static network, i.e. a network which is fixed though perhaps unknown. Departing from prior work, our study analyzes the scenario where *the observed edges which characterize the interference are themselves dependent on the treatment assignments*. This introduces a unique challenge, as the very structure of interference becomes treatment-dependent.

PseudoInverse Estimators Network interference is also related to a problem in slate and combinatorial bandits [Jia et al., 2024, Xu et al., 2024]. Several works have addressed this challenge by assuming specific parametric models, such as linear relationships, to link slate features to outcomes [Auer, 2002, Chu et al., 2011, Qin et al., 2014]. A valuable tool in these settings is the Pseudoinverse estimation Cesa-Bianchi and Lugosi [2012], Rusmevichientong and Tsitsiklis [2010], Dani et al. [2007]. Other studies adopt a similar assumption but operate under a semi-bandit feedback model [Kale et al., 2010, Kveton et al., 2015, Krishnamurthy et al., 2015]. Our solution inspires from these pseudoinverse estimators, but is solving a fundamentally different problem, as the problem of treatment dependent interference is not directly addressable by these methods.

3 NOTATION

We are given a population of n units. Let $\boldsymbol{Z}$ be the treatment assignment vector of the entire population and let $\mathcal{Z}$ denote the treatments' space, e.g., for binary treatments $\mathcal{Z} = \{0,1\}^n$ (see Figure 1). We use the Neyman potential outcome framework [Neyman, 1923, Rubin, 1974], and denote by $Y_i(\boldsymbol{z})$ the potential outcome for each $\boldsymbol{z} \in \mathcal{Z}$. We make observations at unit level and denote these observations as Y_i for unit i.

We will consider randomized Bernoulli designs, i.e., each unit i gets allotted the treatment $z_i = 1$ independently with probability $p_i \in (0,1)$. This is natural and easy to imple-

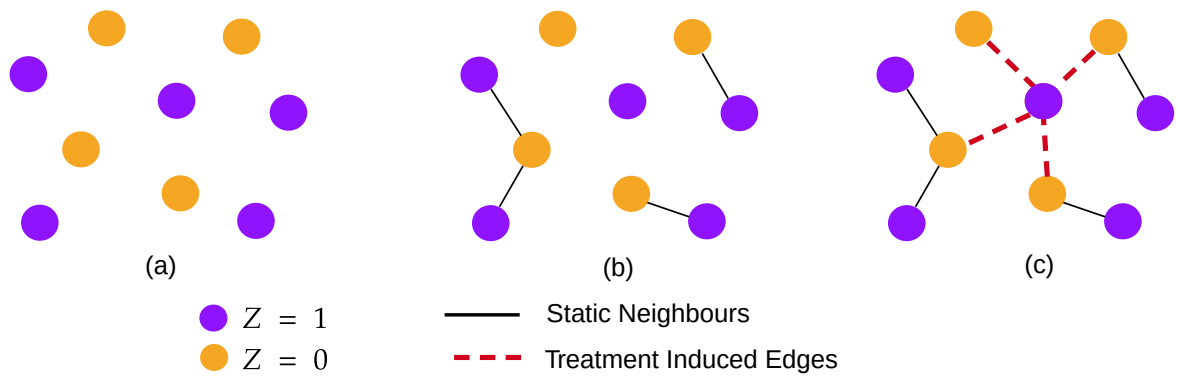


Figure 1: $Z = 1$ denotes the units in the treatment group and $Z = 0$ denotes units in the control group. **(a)** Standard A/B testing where there is no interaction between the treatment and the control units.**(b)** Network interference due to fixed (static) interaction between the units. **(c)** We have network interference however all the red edges are potential edges, and only occured due to specific treatment allocation.

Table 1: Literature Summary. We list a few important works, a few desiderata and whether they are met ✓ or not ×. Our work focuses on the problem of treatment dependent interference which the other methods do not handle.

	General Graph	Uncertain Edges	Single Trial	Treatment Dependent Network
[Hudgens and Halloran, 2008, Liu and Hudgens, 2014]	×	✓	✓	×
[Yuan et al., 2022, Yu et al., 2022]	✓	×	✓	×
[Cortez et al., 2022, Shankar et al., 2023b]	✓	✓	×	×
[Aronow et al., 2017, Sävje et al., 2021b, Toulis and Kao, 2013]	✓	×	✓	×
Ours (Section 5.4.1)	✓	✓	✓	✓
Ours (Section 5.4.2)	✓	✓	✓	✓

ment and satisfies standard randomization and positivity assumptions in causal inference.

Standard Causal Assumptions

Positivity: $P(\boldsymbol{z}) > 0 \, \forall \boldsymbol{z}$ **(A1)**

Consistency: $Y_i = Y_i(\boldsymbol{z})$ if $\boldsymbol{Z} = \boldsymbol{z}$ **(A2)**

We assume that the unit outcome is not determined just by the treatment at the unit but potentially also by treatments allocated to other units. This is a violation of the SUTVA assumption [Cox, 1958, Hudgens and Halloran, 2008] and is commonly called interference.

This dependence can be represented as a graph (Figure 1b), where each node represents a unit and the presence of an edge indicates a possible influence between each other. The underlying graph is given by its adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, with $A_{ij} = 1$ only if an edge exists between from unit j to unit i, and by convention $A_{ii} = 1$. Let $\mathcal{N}_i = \{j : A_{ij} = 1\}$ be the set of *neighbours* of unit i in the unit-unit graph. We assume that the outcomes depend only on the node's neighbours in the unit-unit graph. This is similar to the classic network neighbourhood interference assumption [Hudgens and Halloran, 2008, Sussman and Airoldi, 2017]. However, the classic network interference is not a valid assumption in the scenario we are considering.

Instead, we have a two-stage generative process. We first have a treatment-dependent network formation. Next, conditioned on the network thus formed, the standard network interference assumption is assumed to be valid. To model the network-dependent behaviour, we consider the variables $A_{ij}(\boldsymbol{z})$ as an additional set of potential outcome variables for each possible edge in the network. Corresponding to the potential network edges, we also have neighbourhoods $\mathcal{N}_i(\boldsymbol{z})$. The fundamental interference assumption in our case can be stated as:

Treatment Dependent Network Interference

$$\forall \boldsymbol{z}, \boldsymbol{z}' \text{ s.t. } z_i = z'_i \text{ and } \mathcal{N}_i(\boldsymbol{z}) = \mathcal{N}_i(\boldsymbol{z}') \text{ and } z_j = z'_j \; \forall j \in \mathcal{N}_i(\boldsymbol{z}) : Y_i(\boldsymbol{z}) = Y_i(\boldsymbol{z}'). \quad \textbf{(A3)}$$

Our primary focus is on estimating the Global Average Treatment Effect (GATE) under the previously outlined scenario, where the network structure itself may change based on the chosen treatments. The desired causal effect is the mean difference between the outcomes when $\boldsymbol{z} = \vec{1}\, i.e.,\ z_i = 1\, \forall i$ and when $\boldsymbol{z} = \vec{0}\, i.e.,\ z_i = 0\, \forall i$. Under the aforementioned notations, this causal effect is given by:

$$\tau(\vec{1}, \vec{0}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[Y_i(\vec{1}) - Y_i(\vec{0})] \tag{1}$$

where the expectation $\mathbb{E}$ marginalizes over the different networks. Correspondingly we can also define the individual global treatment effect $\tau_i = \mathbb{E}[Y_i(\vec{1}) - Y_i(\vec{0})]$

SUTVA Estimate The SUTVA estimate (or the DM estimate) is given by

$$\hat{\tau}_{\text{SUTVA}} = \bar{Y}^1 - \bar{Y}^0 = \frac{\sum Y_i \mathbb{I}[Z_i = 1]}{\sum \mathbb{I}[Z_i = 1]} - \frac{\sum Y_i \mathbb{I}[Z_i = 0]}{\sum \mathbb{I}[Z_i = 0]}$$

where $\bar{Y}^{0/1}$ are the average of observed outcomes for units where $Z_i = 0/1$ respectively. This estimator, while simple and practical, requires the SUTVA assumption, and hence can be misleading in our scenario.

4 CHALLENGE AND FORMULATION

Inverse Propensity/Horvitz-Thompson Estimate A classic method to estimate treatment effects is the Horvitz Thompson estimator [Horvitz and Thompson, 1952] (also called IPW or IS estimator). When all treatment decisions are independent Bernoulli variables with probability p_i, the Horvitz Thompson (HT) estimator as follows:

$$\begin{aligned} \tau_{\text{HT}} &= \frac{1}{n} \sum_i Y_i \left(\frac{\prod_{j \in \mathcal{N}_i} z_j}{\prod_{j \in \mathcal{N}_i} p_j} - \frac{\prod_{j \in \mathcal{N}_i} (1 - z_j)}{\prod_{j \in \mathcal{N}_i} (1 - p_j)} \right) \\ &= \frac{1}{n} \sum_i Y_i \left(\prod_{j \in \mathcal{N}_i} \frac{z_j}{p_j} - \prod_{j \in \mathcal{N}_i} \frac{(1 - z_j)}{(1 - p_j)} \right) \end{aligned} \tag{2}$$

If the network is fixed the IPW estimate (and its variants) do not require any further assumption other than randomization and positivity. Unfortunately, when the network is dependent on the treatment vector Z, the HT estimator is not unbiased.

For example, consider 3 node graph with nodes L, R, and U. Each node is a binary treatment node (can be only 0 or 1) (shown in Figure 2)

Edge UL exists if and only if $Z_U = 1$ otherwise the edge UR will exist. However outcomes at L and R, i.e. (Y_L, Y_R) respectively are independent of treatment at U and only depend on treatment at self with the effect being constant

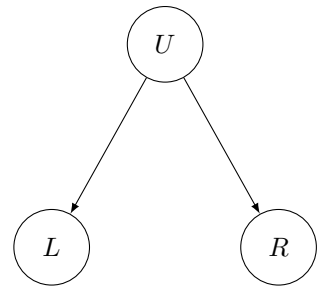


Figure 2: Counterexample demonstrating bias of the standard HT estimate. The figure shows two edges one between U and L, and another between U and R. However, these are potential edges and when treatment allocation happens, only one of the edges will be observed while the other will vanish. The shifting of the edge between counterfactuals causes the bias in HT estimate.

α i.e. the outcomes are $Y_{L/R}(1) = Y_{L/R}(0) + \alpha_{L/R}$. All treatments are randomized with probability $q = 0.5$.

We consider the total treatment effect (TTE) or global average treatment effect (GATE) between $Z = \vec{0}$ and $Z = \vec{1}$ with the HT estimate here. By symmetry we can consider only U, L with the U, R case analogous. Consider the standard HT estimator: we have 4 possibilities for the relevant treatments each with probability 0.25. When $Z_U = 0$, the observed network and counterfactual network is the same; and hence the value of the τ_{HT} is unbiased (= $Y_L(1) - Y_L(0)$. However when $Z_U = 1$, the HT estimator takes into account the edge UL. Thus when $Z_U = 1, Z_L = 0$, the propensity terms in the estimator zero out, leading to 0 value. Thus the expected value of the HT estimator from node L over all treatment allocations is given by $(Y_L(1) + \frac{Y_L(1) - Y_L(0)}{2})$. [1]. Similarly, the contribution from node R is $(-Y_R(0) + \frac{Y_R(1) - Y_R(0)}{2})$.

Hence, the expected value of the estimator for all nodes together is given by $\frac{Y_L(1) - Y_R(0)}{2} + \frac{\alpha_L + \alpha_R}{4}$. On the other hand, the true treatment effect is, the mean of $Y(1) - Y(0)$ over all nodes i.e. $(\alpha_L + \alpha_R)/2$. Thus, we can see that the HT estimator is biased.

The problem arose because if one does not observe the edge between the nodes (L/R) and U, the HT estimator does not include it in the inverse probability weights (since they are dependent on the network) ratio. And between the 2 possibilities the weight ratio moved from L to R (because the edge moved from L to R) in the HT estimate, which caused the bias. We discuss more formally the issue with HT estimation in the Appendix.

Outcome Model (Additive Interference):

$$Y_i(Z) = b_i + c_{ii} Z_i + \sum c_{ij} \tilde{Z}_{ij}$$

[1] More detailed case analysis is in the Appendix

where b_i is the baseline effect, c_{ii} is the direct effect of treatment, $\tilde{Z}_{ij}$ refers to individual factors arising from the treatment vector, and $c_{i,j}$ is the influence of factor j on node i. In the case of standard linear network interference $\tilde{Z}_{ij} = Z_j$. Higher order network dependence can also be modeled here by having multiplicative interaction terms between the components of Z, but for this paper we will focus on the linear case.

Linear Additive Interference

$$\forall i, Y_i(\boldsymbol{z}) = b_i + c_{ii} z_i + \sum c_{ij} A_{ij} z_j \qquad \textbf{(A4)}$$

Remark 4.1. We have not yet assumed anything about $c_{i,j}$, and thus our method supports heterogeneous effects.

Remark 4.2. The presented counter-example presented earlier does satisfy an additive interference. Thus this specific assumption is not enough to solve the problem.

The GATE is defined as: $\tau = \mathbb{E}[Y_i(\vec{1})] - \mathbb{E}[Y_i(\vec{0})]$, where $\vec{1}$ and $\vec{0}$ represent the all 1 (all treated) and all 0 (all untreated) treatment vectors. Substituting this in the outcome model we get

$$\tau_i = c_{ii} + \sum_j c_{ij} \mathbb{E}[A_{ij} | \boldsymbol{z} = \vec{1}]$$

5 ESTIMATION

In this section we first present a general matrix representation framework to estimate the treatment effect τ based on matrix pseudoinverses. We then show how this design fails in the treatment dependent network case, because of a hidden endogeneity. We next discuss how this suggests a solution to the problem by introducing instrument variables.

5.1 MATRIX REPRESENTATION

The discussion in this section follows the presentation of Cesa-Bianchi and Lugosi [2012] Let $\mathcal{N}_i$ be the *fixed* set of neighbours of a specific ego node i. Consider a hypothetical scenario, where we observe a collection of r experiments, each time conducted with a different vector Z. Let Y_i^r be the observed outcome at node i in the r-th trial. Under the linear-additive assumption, we can write:

$$Y_i^r = b_i + c_{ii} + \left(Z^r_{\mathcal{N}(i)}\right)^\top c_i,$$

where $Z^r_{\mathcal{N}(i)}$ is the vector of treatments corresponding to the neighbors of i (or nodes from which i receives interference) in trial r, c_{ii} is the direct effect of treating i, and c_i is the vector of marginal effects of each neighbor's treatment on i. We can formally express the variables from these hypothetical trials as in matrix form as follows:

$$\underbrace{\begin{bmatrix} Y_i^1 \\ Y_i^2 \\ \vdots \\ Y_i^r \end{bmatrix}}_{r\times 1} = \underbrace{\begin{bmatrix} 1 & \left(Z^1_{\mathcal{N}(i)}\right)^\top \\ 1 & \left(Z^2_{\mathcal{N}(i)}\right)^\top \\ \vdots & \vdots \\ 1 & \left(Z^r_{\mathcal{N}(i)}\right)^\top \end{bmatrix}}_{r\times d} \underbrace{\begin{bmatrix} b_i \\ c_{ii} \\ \vec{c}_i \end{bmatrix}}_{d\times 1} \quad \Rightarrow \quad \boldsymbol{Y}_i = \boldsymbol{Z}_i \, \boldsymbol{c}_i.$$

Here, d is the dimension of the parameter vector $\boldsymbol{c}_i$, which includes the direct treatment effect c_{ii} and the vector of neighbor-treatment effects c_i. If we have results from many such random assignments of $\boldsymbol{z}$ make the least square estimator unbiased for $\boldsymbol{c}$.

5.2 TREATMENT *DEPENDENT* GRAPH

:

Now in our scenario, where the network edges depended on treatment allocation, the network structure may change from trial to trial. Consequently, for each experiment r, the set of neighbors $\mathcal{N}(i)$ can vary, leading to different observed components in $Z^r_{\mathcal{N}(i)}$.

Hence, we need to modify the previous approach to include the the variables A_{ij}. We consider the situation in which the node j has an effect on i depends only on z_j, that is, $A_{ij}(\mathbf{z}) = A_{ij}(z_j)$.

The structural equation becomes

$$Y_i = \sum_{j=1}^{n} A_{ij}(Z_j)\, c_{ij}\, Z_j + c_i Z_i. \tag{3}$$

Define the *ideal* (but unobserved) regressors $X_{ij} := A_{ij}(Z_j)\, Z_j$. we have the relation $\boldsymbol{Y}_i = \boldsymbol{X}_i \, \boldsymbol{c}_i$ Once again if we have sufficient number of trials this can be estimated, however that is not feasible in a standard RCT.

With limited number of trials, one cannot observe all the network configurations. Instead one uses the Z based on the network observed in the trial, but the corresponding design matrix ignores the 'counterfactual' edges under alternate treatment allocation.

For simplicity consider the network as obtained from a single trial with the treatment allocation being Z^1. If we naively regress using Z_j from the observed network, then we have

$$X_{ij} = A_{ij}(Z_j^1) Z_j + q_{ij}, \quad q_{ij} := (A_{ij}(Z_j) - A_{ij}(Z_j^1))\, Z_j.$$

Hence the observed design matrix is $W = Z = X - Q$ with $Q = [q_{ij}]$, and (3) can be rewritten as

$$Y_i = \sum_j c_{ij} Z_j + \underbrace{\sum_j c_{ij}(A_{ij}(Z_j) - A_{ij}(Z_j^1)) Z_j}_{\varepsilon_i}.$$

Because ε_i contains functions of Z_j,

$$\mathbb{E}\big[W^\top \boldsymbol{\varepsilon}\big] = \mathbb{E}\big[Z_j\, c_{ij}(A_{ij}(Z_j) - A_{ij}(Z_j^1))Z_j\big] \neq 0.$$

Thus the standard regression assumption of *orthogonality fails*: Z_j is correlated with the regression error, just as in the standard error–in–variables or endogenous regressor problem. Thus if we attempt to apply "static" network interference methods (which assume a fixed set of neighbors and fully observed edges), we end up effectively estimating a regression with an endogenous error term [Sargan, 1958, Bowden and Turkington, 1990].

The presence of the unobserved or "missing" edges shifts part of the structure into an unobserved confounding term, rendering a naive regression approach potentially biased. As detailed, this is reminiscent of *endogenous error* encountered in classical econometrics: the missing (or unobserved) regressors are subsumed into the error term, potentially violating standard exogeneity assumptions. This connection also hints at a solution: the standard method to address endogeneity in econometrics is to use *instrumental variables (IV)*. We propose a similar approach of using IVs. In the next section, we illustrate how IV based methods yield consistent estimates of the treatment (and spillover) effects despite partial observation of the complete network structure.

5.3 IV BASED ESTIMATION

Suppose we have access to mean zero instrumental variables V. From the outcome model

$$Y_i \;= b_i + \; c_{ii} \; + \; (Z_{\mathcal{N}(i)})^\top c_i,$$

we multiply both sides by V and take expectations:

$$\mathbb{E}[VY_i] \;=\; \mathbb{E}[V], c_{ii} + \mathbb{E}[V\,(Z_{\mathcal{N}(i)})^\top]c_i.$$

Since $\mathbb{E}[V] = 0$, the term $\mathbb{E}[V\, b_i]$ vanishes. Solving for c_i yields:

$$c_i \;=\; \Big(\mathbb{E}\big[V\,(Z_{\mathcal{N}(i)})^\top\big]\Big)^{-1}\, \mathbb{E}\big[V\, Y_i\big].$$

Hence, provided $\mathbb{E}[V\,(Z_{\mathcal{N}(i)})^\top]$ is invertible, we can recover c_i consistently by using this moment equation.

Single-Sample Estimation While the above equation holds for expected values, one can obtain consistent estimators by using sample version. Suppose we run R experiments indexed by r, observe $\{V^r, Z^r_{\mathcal{N}(i)}, Y^r_i\}$, and form:

$$\hat{c}_i^R \;=\; \Big[\tfrac{1}{R}\sum_{r=1}^{R} V^r\,\big(Z^r_{\mathcal{N}(i)}\big)^\top\Big]^{-1}\Big[\tfrac{1}{R}\sum_{r=1}^{R} V^r\, Y^r_i\Big].$$

By construction, $\hat{c}_i$ is a consistent estimator of c_i. Moreover, if the matrix $\mathbb{E}[V\,(Z_{\mathcal{N}(i)})^\top]$ is known (or can be computed from external information), then even a single experiment r could suffice. In that scenario,

$$\hat{c}_i \;=\; \Big(\mathbb{E}\big[V\,(Z_{\mathcal{N}(i)})^\top\big]\Big)^{-1}\,\big(V\,Y_i\big),$$

and since $V\,Y_i$ is an unbiased estimate of $\mathbb{E}[V\,Y_i]$, $\hat{c}_i$ remains unbiased.

In the above argument, the matrix $\mathbb{E}[V(Z_{\mathcal{N}i})^T]$ was considered invertible. However this in general will not be the case. For a non-invertible matrix one can use the Moore-Penrose pseudo-inverse. If $\mathbb{E}[V(Z_{Ni})^T]$ has full column rank, the estimates remain unbiased. Thus we have the following estimator

$$\hat{c}_i = \mathbb{E}[V(Z_{\mathcal{N}i})^T]^{+}[\sum V^r Y_i^r] \tag{4}$$

5.3.1 Identification Condition

For identification,we require the following conditions

- **Relevance**: V is correlated with $Z_{\mathcal{N}_i}$,
- **Exclusion**: V affects Y_i only through Z_{N_i}.

Both of these conditions are natural in the standard IV literature [Angrist et al., 1996, Sargan, 1958, Bowden and Turkington, 1990, Bonet, 2013]. Relevance ensures that V captures enough variation in Z to ensure $\mathbb{E}[V(Z_{\mathcal{N}i})^T]$ is non singular. Exclusion ensures that VY_i does not have any systematic Z dependent component.

A common instrument in network settings is the treatment of neighbours [Drago et al., 2020, Rogowski and Sinclair, 2012]. In our setting also, these variables can serve as valid instrument variables [Rogowski and Sinclair, 2012]. Specifically, we will use for each node j we can create an instrument $V_j = \frac{Z_j}{p} - \frac{(1-Z_j)}{(1-p)}$. By construction, V_j it is correlated with the $Z_{\mathcal{N}(i)}$ if $j \in \mathcal{N}(i)$, thus satisfying relevance. However, exclusion is not always satisfied, specifically if j appears in $\mathcal{N}(i)$ for one allocation but not in a different one. Next, we describe detail two specific methods leveraging the aforementioned idea of IV based pseudoinverse estimator, by using two different constructions of neighbourhood based IVs.

5.4 ESTIMATORS

5.4.1 Overcomplete Estimator

Consider the scenario, when for each node i we know a superset of all possible neighbours under all possible treatment allocations. Lets denote this set as $\mathcal{M}_i$.

Neighbourhood Superset: $\mathcal{M}_i \supseteq \mathcal{N}_i(\boldsymbol{z})\ \forall i, \boldsymbol{z}$ **(A5)**

In such a case, the treatment of all units in $\mathcal{M}$ provides an overcomplete set of instruments.

The estimator is present in Equation (5)In this setting, the GATE estimator becomes the estimator of Sussman and Airoldi [2017], which itself can be seen as a version of the standard pseudo-inverse estimator [Swaminathan et al., 2017, Cesa-Bianchi and Lugosi, 2012].

$$\hat{\tau}_{\text{OIV}} = \frac{1}{n}\sum_i Y_i \sum_{j\in\mathcal{M}_i}\left(\frac{z_j}{p} - \frac{(1-z_j)}{(1-p)}\right). \tag{5}$$

The derivation of the above estimator from Equation (4) is in the Appendix (Lemma A.5).

Proposition 5.1. *Under assumptions **A1-4,A5** , $\hat{\tau}_{OIV}$ is an unbiased estimate of the treatment effect τ*

Remark 5.2. While Assumption **A5** can be a strong assumption, in many scenarios this can be satisfied. As a simple example, consider all nodes which share a geographic location (or in case of units being mobile devices, IP). This is very likely to be a superset of all interactions this unit can have. In other cases, user modeling and device-linking methods are used to identify neighbours based on confidence scores i.e. they have a probabilistic version of the adjacency matrix $\boldsymbol{A}$. Such a method can usually be adapted to obtain a superset of neighbours with high probability (by including even low confidence nodes as neighbours.

We now turn to the case when we do not have enough IVs. For the linear case we would have required as many instruments as nodes. This along with the relevance criteria can be hard to satisfy, and so a method which works with fewer instruments is more valuable for some applications.

5.4.2 Undercomplete Estimator

In this section we consider the case of undercomplete V. As earlier the treatment of neighbouring nodes are used to create the instrument. However, the set of observed neighbours do not qualify as valid instruments [2]. The method from the previous section used a superset $\mathcal{M}_i$ of all possible neighbours; or equivalently a set which is the union of all the neighbouring sets under all possible treatments.

Now we present an alternative which instead relies on the intersection of all the neighbouring sets under all possible treatments. Equivalently consider the set of edges $j \to i$ such that $A_{ij}(\boldsymbol{Z})$ is a constant function independent of Z.

[2]Using only the observed neighbours is the same as assuming static interference, which as shown earlier leads to biased estimation

These set of edges will continue to exist regardless of treatment assignments, and thus we call them conserved edges. Let use denote such a set of edges as $\mathcal{M}_i^c$ The knowledge of a large enough set of pre-experiment edges that are conserved, allows us to circumvent the difficulties posed by not observing edges under counterfactual treatments.

Conserved Set: $\mathcal{M}_i^c \subseteq \mathcal{N}_i(\boldsymbol{z})\ \forall i, \boldsymbol{z}$ **(A6)**

Remark 5.3. The existence of such a conserved edges is analogous to the classical "compliance" assumption used for instrumental variables estimation[Angrist et al., 1996].

We propose to use the IV pseudo-inverse estimator 4, but will adjust the estimate obtained, by noting that it only covers a subset of the variables. Such a set is almost always by construction undercomplete. However we also note that we do not need the entire vector c_i. Instead we care only about the total treatment effect which is $c_i^T\vec{1}$. Under certain assumptions, the estimate obtained by using the undercomplete pseudo-inverse can be adjusted to be unbiased.

One such assumption is the assumption of homogenous neighbours (**A7**). Under this assumption c_{ij} does not depend on j. Hence, this is also called anonymous interference as the effect does not depend on the identity of the neighour.

Anonymous Interference: $c_{ij} = c_{ij'} \forall j, j' \in \mathcal{N}_i \setminus i$ **(A7)**

Remark 5.4. c_{ij} can still depend on i, so we still have some heterogeneity.

Let $C_i = \frac{1}{p}\sum_j Z_j A_{ij}$, then

$$\mathbb{E}[C_i] = \sum_j \frac{1}{p}\mathbb{E}[Z_j A_{ij}] = \sum_j \mathbb{E}[A_{ij}|Z_j = 1] \tag{6}$$

One key result that (see the Appendix) is that, if we use $\mathcal{M}_i^c$ as the instrument, the pseudo-inverse provides an unbiased estimate of the indirect effect of nodes in $\mathcal{M}_i^c$. That is we have $\mathbb{E}[\sum \hat{c}_i] = \sum_{j\in\mathcal{M}_i^c} c_{ij}\mathbb{E}[A_{ij}(1)]$ which under anonymity is just $c_i \sum_{j\in\mathcal{M}_i^c}\mathbb{E}[A_{ij}(1)]$ which further under conserved edges becomes $c_i|\mathcal{M}_i^c|$. Thus we can rescale this estimate by C_i to get an unbiased estimate of τ_i.

$$\hat{\tau}_{\text{UIV}} = \frac{1}{n}\sum_i Y_i\Bigg[\left(\frac{z_i}{p} - \frac{1-z_i}{1-p}\right) + \tag{7}$$

$$\sum_{j\in\mathcal{M}_i^c}\left(\frac{z_j}{p} - \frac{(1-z_j)}{(1-p)}\right)\left(\frac{\sum_j z_j}{p|\mathcal{M}_i^c|}\right)\Bigg]. \tag{8}$$

Proposition 5.5. *Under assumptions **A1-4,A6-7** , $\hat{\tau}_{UIV}$ is an unbiased estimate of the treatment effect τ*

We would like to bring a crucial detail to the attention of the reader. As mentioned before $\hat{\tau}_{OIV}$ is very similar to the HATE estimator of Sussman and Airoldi [2017]. Similarly $\hat{\tau}_{UIV}$ is a scaled version of the same estimator. The critical difference between them lies in the set of neighbours used. This is because under treatment dependent networks, the neighbourhood itself also becomes a function of treatment, and using the observed neighbourhood will cause errors. How $\hat{\tau}_{OIV}$,$\hat{\tau}_{UIV}$ specifically handle this is discussed in more detail in Appendix A.2.

Remark 5.6. In Appendix A.1, we derive bounds for the variance of the UIV and OIV estimator which can be used to provide conservative intervals for a Wald-style hypothesis test [Wasserman, 2006].

Remark 5.7. We present another estimator based on the insight from (Equation (6)) in the Appendix. This estimator, while efficient and with quite low variance, requires multiple independent trials. Due to these conditions, this estimator is not applicable for many real datasets where we conduct the experiment once. That said for certain applications, researchers have access to baseline results [Cortez et al., 2022] which can be used as a trial.

6 EXPERIMENTS

6.1 SYNTHETIC GRAPHS

In this section, we experimentally demonstrate the validity of our proposed methods by experimenting with synthetic data obtained from a model which satisfies our assumptions exactly. We experiment with both Erdős-Rényi (ER) graphs and stochastic block model (SBM) graphs to compare the performance of our estimator with other estimators. We simulate 100 different random graphs and run repeated experiments on each graph with random treatment assignments. We set a independence parameter e which determines the fraction of these edges which will not show a treatment dependent behaviour. Specifically each treatment dependent edge acts as a bernoulli variable and will be activated if its source node has treatment 1. A subset of the non-varying neighbourhood is take as the conserved edges for ($\mathcal{M}_i^c$). On the other hand the base network itself is taken to be the superset neighbourhood ($\mathcal{M}_i$). The potential outcomes $Y_i(\boldsymbol{z})$ are obtained by applying a function g on the exposure and adding a mean zero noise. The exposure are computed using the procedure in Cortez et al. [2022]. For each experiment, we varied the treatment probability p, the size of the graphs n to assess the efficacy of estimation across different ranges of parameters and the strength of interference r. Similar to Cortez et al. [2022] we measure the strength of interference r as the ratio of norms of the self or direct influence and the indirect influence (more details in Appendix C.1).

We gauge the effectiveness of MEX by benchmarking it against commonly employed estimators such as polynomial regression (Poly), ReFeX [Han and Ugander, 2023], and the difference-in-means (DM) estimators ($\hat{\tau}_{\text{SUTVA}}$). Due to the size of neighbourhoods, Horwitz-Thompson estimators failed to yield non-meaningful results in these trials.

The results are presented in Figure 3. The first row contains results from the ER model. From the figure it is clear that our model produces unbiased estimates. On the other hand, all other methods produce highly biased estimates. Note that in Figure 3a, when $r = 0$, there is no interference, and hence most estimators are unbiased. However, when interference increases these methods clearly show strong bias. Secondly, for a given interference strength, our method shows consistency in the form of decreasing variance with increasing number of nodes. Finally we also show bias due to treatment dependence in all methods, while we remains unbiased. Similar results are obtained on the SBM model as well.

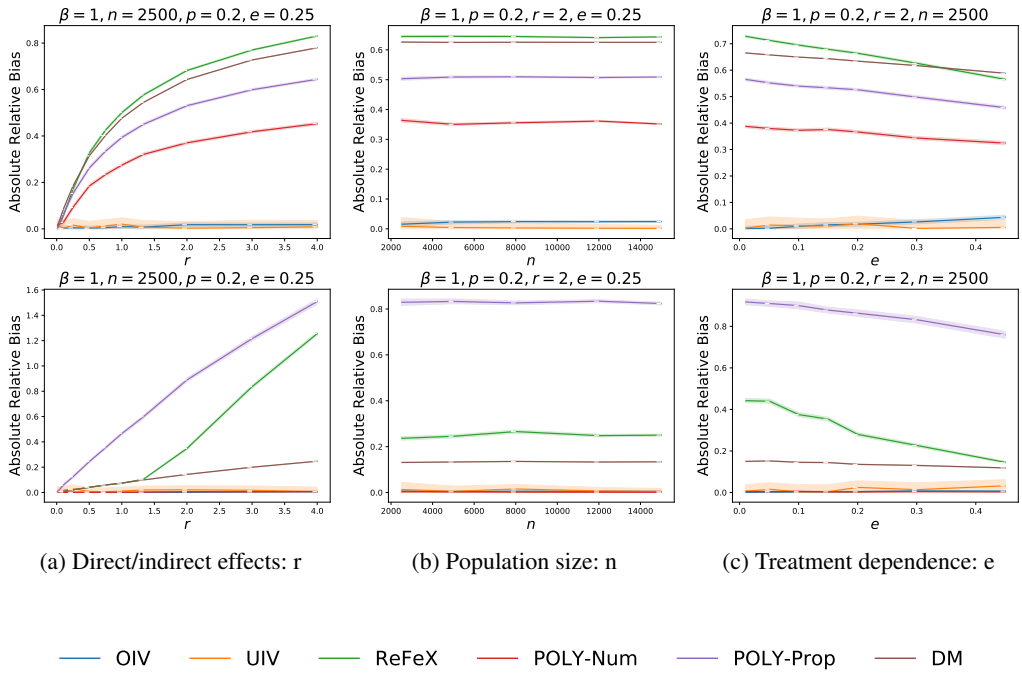
6.2 APPLICATION: ASSESSING IMPACT OF BANKING ACCESS INTERVENTION

Next, we demonstrate an application of observational data. We focus on the application mentioned in the introduction, which introduces access to financial accounts. We use the data from the field study conducted by Prina [2015], Comola and Prina [2021] in the region around Pokhara in Nepal. The experiment involves a randomized trial of providing access to formal savings accounts to a random sample of poor households. The authors surveyed all poor households with an adult and working female head to identify their social connections. The initial social network was sparse and minimally clustered. Half of the families were offered access to a savings bank account. After the treatment, another survey was conducted with the families. Comola and Prina [2021] have reported a significant fraction of treated units using the savings bank account. They also found a significant change in social connections, with around 50% of the connections changing post-treatment. The outcomes Y_i correspond to measured household consumption. Literature has shown strong peer effects for this variable [Cruwys et al., 2015]. We used the intersection of the two networks as $\mathcal{M}^c$ for the UIV estimate and their union as $\mathcal{M}$ for the OIV estimate.

As this is observational data, we do not know the ground truth effect and consider the results of Comola and Prina [2021] as a reference. Figure 4 shows that our method provides similar estimates as the the reference, but other interference aware methods like RefeX method, while better than no-interference model do not do as well.

7 CONCLUSION

We presented a major limitation of current interference-aware GATE methods. We show that the standard HT esti-



(a) Direct/indirect effects: r (b) Population size: n (c) Treatment dependence: e

Figure 3: Plots visualizing the performance of various GATE estimators under Bernoulli design on Erdős-Rényi networks (first row) and SBM networks (second row). The lines represent the empirical relative bias, i.e., $\frac{\hat{\tau}-\tau}{\tau}$ of the estimators across different settings, with the shaded width corresponding to the experimental standard error.

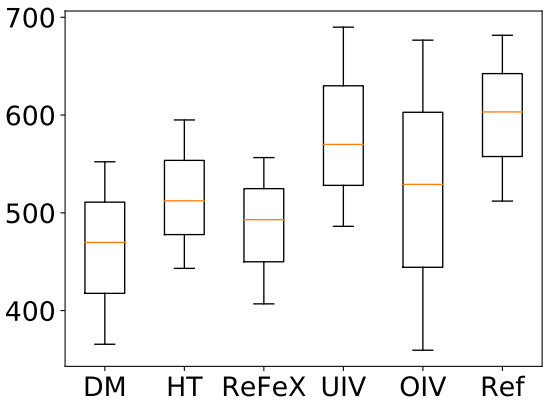


Figure 4: Estimates for GATE of financial access on household consumption for the [Prina, 2015] experiment. The box plot depicts the mean and the 95% confidence interval. HT and ReFeX methods use post-treatment neighbourhoods, and Ref is the method from Comola and Prina [2021]

mate is biased when the interference network is treatment-dependent. We then provide two different solutions to this problem by combining the ideas of pseudoinverse estimation with the concept of instrumental variables. We show that our estimators are unbiased and provide a statistical inference method. Finally, we experiment with both real and synthetic data to show the validity of our estimators. Our results have immediate implications for randomized trials in social networks, public health, and economics, where ignoring endogenous interference can lead to severely misleading conclusions.

A limitation of our work is that the variance of the estimate grows with the size of the neighbourhoods, and so for practical applications, one needs to balance the risk of higher variance against potential bias. Future research directions include incorporating temporal data and longitudinal studies.

References

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91 (434):444–455, 1996.

Elliott M Antman, Joseph Lau, Bruce Kupelnick, Frederick Mosteller, and Thomas C Chalmers. 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, 1992.

Sinan Aral and Dylan Walker. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57(9):1623–1639, 2011.

Ilkka Arminen. Therapeutic interaction. In *A study of mutual help in the meetings of Alcoholics Anonymous*, volume 45, 1998.

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

Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

Eric Auerbach and Max Tabord-Meehan. The local approach to causal inference under network interference. *arXiv preprint arXiv:2105.03810*, 2021.

Eric Auerbach, Jonathan Auerbach, and Max Tabord-Meehan. Exposure effects are policy relevant only under strong assumptions about the interference structure. *arXiv preprint arXiv:2401.06264*, 2024.

Eytan Bakshy, Dean Eckles, Rong Yan, and Itamar Rosenn. Social influence in social advertising: evidence from field experiments. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 146–161, 2012.

Falco J Bargagli-Stoffi, Costanza Tortù, and Laura Forastiere. Heterogeneous treatment and spillover effects under clustered network interference. *arXiv preprint arXiv:2008.00707*, 2020.

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

Rohit Bhattacharya, Daniel Malinsky, and Ilya Shpitser. Causal inference under interference and network uncertainty. In Ryan P. Adams and Vibhav Gogate, editors, *Proceedings of the 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pages 1028–1038, 22–25 Jul 2020. URL `https://proceedings.mlr.press/v115/bhattacharya20a.html`.

Blai Bonet. Instrumentality tests revisited. *arXiv preprint arXiv:1301.2258*, 2013.

Roger J Bowden and Darrell A Turkington. *Instrumental variables*. Number 8. Cambridge university press, 1990.

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

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

Raymond J Carroll, David Ruppert, Leonard A Stefanski, and Ciprian M Crainiceanu. *Measurement error in nonlinear models: a modern perspective*. CRC Press, 2006.

Nicolo Cesa-Bianchi and Gábor Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78 (5):1404–1422, 2012.

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

Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 208–214. JMLR Workshop and Conference Proceedings, 2011.

Margherita Comola and Silvia Prina. Treatment effect accounting for network changes: Evidence from a randomized intervention. *Available at SSRN 2250748*, 2021.

Ilja Cornelisz, Pim Cuijpers, Tara Donker, and Chris van Klaveren. Addressing missing data in randomized clinical trials: A causal inference perspective. *PloS one*, 15(7): e0234349, 2020.

Mayleen Cortez, Matthew Eichhorn, and Christina Lee Yu. Graph agnostic estimators with staggered rollout designs under network interference. *Advances in Neural Information Processing Systems*, 35:7437–7449, 2022.

Mayleen Cortez-Rodriguez, Matthew Eichhorn, and Christina Lee Yu. Exploiting neighborhood interference with low order interactions under unit randomized design. *Journal of Causal Inference*, 11(1), 2023.

David Roxbee Cox. *Planning of experiments.* Wiley, 1958.

Tegan Cruwys, Kirsten E Bevelander, and Roel CJ Hermans. Social modeling of eating: A review of when and why social influence affects food intake and choice. *Appetite*, 86:3–18, 2015.

Yifan Cui, Hongming Pu, Xu Shi, Wang Miao, and Eric Tchetgen Tchetgen. Semiparametric proximal causal inference. *Journal of the American Statistical Association*, 119(546):1348–1359, 2024.

Varsha Dani, Sham M Kakade, and Thomas Hayes. The price of bandit information for online optimization. *Advances in Neural Information Processing Systems*, 20, 2007.

Vincent Dorie, Masataka Harada, Nicole Bohme Carnegie, and Jennifer Hill. A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. *Statistics in Medicine*, 35(20):3453–3470, 2016.

Francesco Drago, Friederike Mengel, and Christian Traxler. Compliance behavior in networks: Evidence from a field experiment. *American Economic Journal: Applied Economics*, 12(2):96–133, 2020.

Oliver Dukes and Stijn Vansteelandt. Inference for treatment effect parameters in potentially misspecified high-dimensional models. *Biometrika*, 108(2):321–334, 2021.

Oliver Dukes, Ilya Shpitser, and Eric J Tchetgen Tchetgen. Proximal mediation analysis. *Biometrika*, 110(4):973–987, 2023.

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

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

Kenneth A Frank and Ran Xu. Causal inference for social network analysis. *The Oxford handbook of social networks*, pages 288–310, 2020.

Mengsi Gao. Endogenous interference in randomized experiments. *arXiv preprint arXiv:2412.02183*, 2024.

Huan Gui, Ya Xu, Anmol Bhasin, and Jiawei Han. Network A/B testing: From sampling to estimation. In *Proceedings of the 24th International Conference on World Wide Web*, pages 399–409, 2015.

Wenshuo Guo, Mingzhang Yin, Yixin Wang, and Michael Jordan. Partial identification with noisy covariates: A robust optimization approach. 2022. URL `https://openreview.net/forum?id=-NVBxy0TdU`.

Kevin Han and Johan Ugander. Model-based regression adjustment with model-free covariates for network interference. *Journal of Causal Inference*, 11(1), 2023.

James J Heckman and Rodrigo Pinto. Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric reviews*, 34(1-2):6–31, 2015.

Miquel A Hernán and James M Robins. *Causal inference: What if*. Boca Raton: Chapman & Hall/CRC, 2021.

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

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

Guido W Imbens. Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review*, 93(2): 126–132, 2003.

Su Jia, Peter Frazier, and Nathan Kallus. Multi-armed bandits with interference. *arXiv preprint arXiv:2402.01845*, 2024.

Yiming Jiang and He Wang. Causal inference under network interference using a mixture of randomized experiments. *arXiv preprint arXiv:2309.00141*, 2023.

Satyen Kale, Lev Reyzin, and Robert E Schapire. Non-stochastic bandit slate problems. *Advances in Neural Information Processing Systems*, 23, 2010.

Brian Karrer, Liang Shi, Monica Bhole, Matt Goldman, Tyrone Palmer, Charlie Gelman, Mikael Konutgan, and Feng Sun. Network experimentation at scale. In *Proceedings of the 27th acm sigkdd conference on knowledge discovery & data mining*, pages 3106–3116, 2021.

Noémi Kreif, Susan Gruber, Rosalba Radice, Richard Grieve, and Jasjeet S Sekhon. Evaluating treatment effectiveness under model misspecification: a comparison of targeted maximum likelihood estimation with bias-corrected matching. *Statistical methods in medical research*, 25(5):2315–2336, 2016.

Akshay Krishnamurthy, Alekh Agarwal, and Miroslav Dudik. Efficient contextual semi-bandit learning. *arXiv preprint arXiv:1502.05890*, 2015.

Manabu Kuroki and Judea Pearl. Measurement bias and effect restoration in causal inference. *Biometrika*, 101(2): 423–437, 2014.

Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvari. Tight regret bounds for stochastic combinatorial semi-bandits. In *Artificial Intelligence and Statistics*, pages 535–543. PMLR, 2015.

James LeSage and Robert Kelley Pace. *Introduction to spatial econometrics*. Chapman and Hall/CRC, 2009.

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

Michael P Leung. Causal inference under approximate neighborhood interference. *Econometrica*, 90(1):267–293, 2022.

Wei Li and Xiao-Hua Zhou. Identifiability and estimation of causal mediation effects with missing data. *Statistics in Medicine*, 36(25):3948–3965, 2017.

Wenrui Li, Daniel L Sussman, and Eric D Kolaczyk. Causal inference under network interference with noise. *arXiv preprint arXiv:2105.04518*, 2021.

Lan Liu and Michael G. Hudgens. Large sample randomization inference of causal effects in the presence of interference. *Journal of the American Statistical Association*, 109(505):288–301, 2014. doi: 10.1080/01621459.2013.844698. URL `https://doi.org/10.1080/01621459.2013.844698`. PMID: 24659836.

JR Lockwood and Daniel F McCaffrey. Matching and weighting with functions of error-prone covariates for causal inference. *Journal of the American Statistical Association*, 111(516):1831–1839, 2016.

Wang Miao, Zhi Geng, and Eric J Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, 2018.

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

Elizabeth L Ogburn and Tyler J Vanderweele. Bias attenuation results for nondifferentially mismeasured ordinal and coarsened confounders. *Biometrika*, 100(1):241–248, 2013.

Elizabeth L Ogburn and Tyler J VanderWeele. Causal diagrams for interference. *Statistical science*, 29(4):559–578, 2014.

Elizabeth L Ogburn, Oleg Sofrygin, Ivan Diaz, and Mark J Van der Laan. Causal inference for social network data. *arXiv preprint arXiv:1705.08527*, 2017.

Elizabeth L Ogburn, Oleg Sofrygin, Ivan Diaz, and Mark J Van der Laan. Causal inference for social network data. *Journal of the American Statistical Association*, 119(545): 597–611, 2024.

Michael O'Riordan and Ciaran M Gilligan-Lee. Local interference bias in semi-parametric models. *arXiv preprint*, 2025.

Judea Pearl. On measurement bias in causal inference. *arXiv preprint arXiv:1203.3504*, 2012.

Jean Pouget-Abadie. *Dealing with Interference on Experimentation Platforms*. PhD thesis, Harvard University, 2018.

Jean Pouget-Abadie, Martin Saveski, Guillaume Saint-Jacques, Weitao Duan, Ya Xu, Souvik Ghosh, and Edoardo Maria Airoldi. Testing for arbitrary interference on experimentation platforms. *arXiv preprint arXiv:1704.01190*, 2017.

Silvia Prina. Banking the poor via savings accounts: Evidence from a field experiment. *Journal of development economics*, 115:16–31, 2015.

Lijing Qin, Shouyuan Chen, and Xiaoyan Zhu. Contextual combinatorial bandit and its application on diversified online recommendation. In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 461–469. SIAM, 2014.

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

Jon C Rogowski and Betsy Sinclair. Estimating the causal effects of social interaction with endogenous networks. *Political Analysis*, 20(3):316–328, 2012.

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

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

Donald B. Rubin. Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics*, 6(1):34 – 58, 1978. doi: 10.1214/aos/1176344064. URL `https://doi.org/10.1214/aos/1176344064`.

Paat Rusmevichientong and John N Tsitsiklis. Linearly parameterized bandits. *Mathematics of Operations Research*, 35(2):395–411, 2010.

Rishiraj Saha Roy, Ritwik Sinha, Niyati Chhaya, and Shiv Saini. Probabilistic deduplication of anonymous web traffic. In *Proceedings of the 24th International Conference on World Wide Web*, 2015.

John D Sargan. The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the econometric society*, pages 393–415, 1958.

Fredrik Sävje. Causal inference with misspecified exposure mappings: separating definitions and assumptions. *Biometrika*, 111(1):1–15, 2024.

Fredrik Sävje, Peter Aronow, and Michael Hudgens. Average treatment effects in the presence of unknown interference. *Annals of statistics*, 49(2):673, 2021a.

Fredrik Sävje, Peter M Aronow, and Michael G Hudgens. Average treatment effects in the presence of unknown interference. *The Annals of Statistics*, 49(2):673–701, 2021b.

Noah A Schuster, Judith JM Rijnhart, Lisa C Bosman, Jos WR Twisk, Thomas Klausch, and Martijn W Heymans. Misspecification of confounder-exposure and confounder-outcome associations leads to bias in effect estimates. *BMC medical research methodology*, 23(1):11, 2023.

Shiv Shankar, Ritwik Sinha, Saayan Mitra, Moumita Sinha, and Madalina Fiterau. 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)*, 2023a.

Shiv Shankar, Ritwik Sinha, Saayan Mitra, Viswanathan (Vishy) Swaminathan, Sridhar Mahadevan, and Moumita Sinha. 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, 2023b.

Shiv Shankar, Ritwik Sinha, Yash Chandak, Saayan Mitra, and Madalina Fiterau. A/b testing under interference with partial network information. In *International Conference on Artificial Intelligence and Statistics*, pages 19–27. PMLR, 2024a.

Shiv Shankar, Ritwik Sinha, and Madalina Fiterau. Estimating counterfactual distributions under interference. In *Proceedings of Machine Learning in Healthcare*. PMLR, 2024b.

Shiv Shankar, Ritwik Sinha, and Madalina Fiterau. On online experimentation without device identifiers. In *Forty-first International Conference on Machine Learning*, 2024c.

Shiv Shankar, Ritwik Sinha, and Madalina Fiterau. Online experimentation under privacy induced identity fragmentation. In *Privacy Regulation and Protection in Machine Learning*, 2025.

Ilya Shpitser, Chan Park, Eric Tchetgen Tchetgen, and Ryan Andrews. Modeling interference via symmetric treatment decomposition. *arXiv preprint arXiv:1709.01050*, 2017.

Ilya Shpitser, Zach Wood-Doughty, and Eric J Tchetgen Tchetgen. The proximal ID algorithm. *arXiv preprint arXiv:2108.06818*, 2021.

Ritwik Sinha, Shiv Saini, and N Anadhavelu. Estimating the incremental effects of interactions for marketing attribution. In *2014 International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC2014)*, pages 1–6. IEEE, 2014.

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

Meta-Lina Spohn, Leonard Henckel, and Marloes H Maathuis. A graphical approach to treatment effect estimation with observational network data. *arXiv preprint arXiv:2312.02717*, 2023.

Daniel L Sussman and Edoardo M Airoldi. Elements of estimation theory for causal effects in the presence of network interference. *arXiv preprint arXiv:1702.03578*, 2017.

Adith Swaminathan, Akshay Krishnamurthy, Alekh Agarwal, Miro Dudik, John Langford, Damien Jose, and Imed Zitouni. Off-policy evaluation for slate recommendation. *Advances in Neural Information Processing Systems*, 30, 2017.

Tracy Sweet and Samrachana Adhikari. A latent space network model for social influence. *Psychometrika*, 85 (2):251–274, 2020.

Eric J Tchetgen Tchetgen and Tyler J VanderWeele. On causal inference in the presence of interference. *Statistical Methods in Medical Research*, 21(1):55–75, 2012. doi: 10.1177/0962280210386779. URL `https://doi.org/10.1177/0962280210386779`. PMID: 21068053.

Eric J Tchetgen Tchetgen, Andrew Ying, Yifan Cui, Xu Shi, and Wang Miao. An introduction to proximal causal learning. *arXiv preprint arXiv:2009.10982*, 2020.

Panos Toulis and Edward Kao. Estimation of causal peer influence effects. In *International Conference on Machine Learning*, pages 1489–1497, 2013.

Linda Valeri and Tyler J Vanderweele. The estimation of direct and indirect causal effects in the presence of misclassified binary mediator. *Biostatistics*, 15(3):498–512, 2014.

Tyler J. VanderWeele, Eric J. Tchetgen Tchetgen, and M. Elizabeth Halloran. Interference and sensitivity analysis. *Statist. Sci.*, 29(4):687–706, 11 2014. doi: 10.1214/14-STS479. URL `https://doi.org/10.1214/14-STS479`.

Stijn Vansteelandt, Maarten Bekaert, and Gerda Claeskens. On model selection and model misspecification in causal inference. *Statistical methods in medical research*, 21(1): 7–30, 2012.

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

Davide Viviano. Experimental design under network interference. *arXiv preprint arXiv:2003.08421*, 2020.

Yingfei Wang, Hua Ouyang, Chu Wang, Jianhui Chen, Tsvetan Asamov, and Yi Chang. Efficient ordered combinatorial bandits for whole-page recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

Yixin Wang, Dawen Liang, Laurent Charlin, and David M Blei. Causal inference for recommender systems. In *Fourteenth ACM Conference on Recommender Systems*, pages 426–431, 2020.

Larry Wasserman. *All of nonparametric statistics*. Springer Science & Business Media, 2006.

Yang Xu, Wenbin Lu, and Rui Song. Linear contextual bandits with interference. *arXiv preprint arXiv:2409.15682*, 2024.

Steve Yadlowsky, Hongseok Namkoong, Sanjay Basu, John Duchi, and Lu Tian. Bounds on the conditional and average treatment effect with unobserved confounding factors. *arXiv preprint arXiv:1808.09521*, 2018.

Grace Y. Yi, Aurore Delaigle, and Paul Gustafson. *Handbook of Measurement Error Models*. CRC Press, 2021.

Mingzhang Yin, Claudia Shi, Yixin Wang, and David M Blei. Conformal sensitivity analysis for individual treatment effects. *arXiv preprint arXiv:2112.03493*, 2021.

Christina Lee Yu, Edoardo Airoldi, Christian Borgs, and Jennifer Chayes. Estimating total treatment effect in randomized experiments with unknown network structure. *Proceedings of the National Academy of Sciences*, 119 (44), 2022.

Yuan Yuan, Kristen Altenburger, and Farshad Kooti. Causal network motifs: identifying heterogeneous spillover effects in A/B tests. In *Proceedings of the Web Conference 2021*, pages 3359–3370, 2022.

Chi Zhang, Karthika Mohan, and Judea Pearl. Causal inference under interference and model uncertainty. In *Proceedings of the Second Conference on Causal Learning and Reasoning*, pages 371–385, 2023.

Junzhe Zhang and Elias Bareinboim. Bounding causal effects on continuous outcomes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12207–12215, 2021.

Qingyuan Zhao, Dylan S Small, and Bhaswar B Bhattacharya. Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. *arXiv preprint arXiv:1711.11286*, 2017.

A RELATED WORK

Network Interference Network interference is a well studied topic in causal inference literature, with a variety of methods proposed for the problem. Existing works in this area incorporate various sets of assumptions to provide an estimate of treatment effects. A common approach is the exposure mapping framework which allows defines a degree of "belonging" of a unit to either the treatment or control group [Aronow et al., 2017, Auerbach and Tabord-Meehan, 2021, Li et al., 2021, Viviano, 2020]. Typically linearity with respect to neighbouring treatments is also assumed [Eckles et al., 2017, Leung, 2022, Zhang et al., 2023, Wang et al., 2017] but is not neccessary [Sussman and Airoldi, 2017]. A limitation of these approaches is that they require complete knowledge of the network structure.

Treatment effect estimation with unknown network interference has been studied beginning with the seminal work of Hudgens and Halloran [2008]. Other works such as Auerbach and Tabord-Meehan [2021], Bhattacharya et al. [2020], Liu and Hudgens [2014], Tchetgen Tchetgen and VanderWeele [2012], VanderWeele et al. [2014] have extended this idea further. Often the bias of these estimators depends on the the number of edges between the clusters, but constructing good clusters is also known to be computationally intractable[Pouget-Abadie, 2018]. This has led to development of various heuristic methods for constructing clusters [Eckles et al., 2017, Gui et al., 2015]. However, this still requires the graph to be static and not treatment dependent. On the other hand, *our method can handle treatment dependence in general unstructured graphs*.

Estimation with Misspecifications and Mismeasurements The estimation of treatment effects in the presence of model misspecification is an important problem in causal inference, with numerous methods and heuristics proposed to address this challenge [Carroll et al., 2006, Ogburn and Vanderweele, 2013, Lockwood and McCaffrey, 2016]. A comprehensive overview on this problem can be found in Yi et al. [2021], Vansteelandt et al. [2012]. Various approaches have been proposed towards handling misspecification in noise model Dukes and Vansteelandt [2021], propensity weights [Kreif et al., 2016], confounders Pearl [2012], Schuster et al. [2023], and mediators Valeri and Vanderweele [2014], Dukes et al. [2023].

A related problem to misspecifed models is noisy measurements. In general access to noisy variables is not sufficient to identify causal effects [Kuroki and Pearl, 2014, Hernán and Robins, 2021]. Some research in solving this problem [Dukes et al., 2023, Cui et al., 2024] uses ideas from proximal causal inference [Tchetgen Tchetgen et al., 2020]. However these require knowledge of multiple . A different approach has been to focus on bounding for treatment effects rather than estimate them precisely. This line of work includes methods for sensitivity analysis [Imbens, 2003, Veitch and Zaveri, 2020, Dorie et al., 2016] and partial identification under various assumptions [Zhao et al., 2017, Yadlowsky et al., 2018, Zhang and Bareinboim, 2021, Yin et al., 2021, Guo et al., 2022]. Similar analysis for missing data has been conducted for missing mediators [Li and Zhou, 2017] and outcomes [Cornelisz et al., 2020]

Existing methods for causal effect estimation under imprecise networks often require additional information to mitigate bias. For example, some approaches leverage repeated measurements to reduce the impact of noise [Shankar et al., 2023b, Cortez et al., 2022], while others rely on a gold standard sample of measurements to calibrate or correct noisy data [Shankar et al., 2023a]. *These strategies however do not apply when the networks are treatment dependent. This is because compared to these earlier works, the noise acts as an endogenous variable, which needs specialized techniques.*

Inverse Propensity/Horvitz-Thompson Estimate If the graph is known and when all treatment decisions independently set with probability p, one can use the classic Horvitz Thompson estimator (or inverse propensity estimator) as:

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

A similar formula exists for the Hajek style estimator with the denominators $\prod_{j\in\mathcal{N}_i} p$ and $\prod_{j\in\mathcal{N}_i}(1-p)$, replaced by their self normalized values. This estimator filters out any units for which all neighbours are not in control or treatment groups, and is not be meaningful, when there do not not exist units for which all the neighbours are in control or treatment groups. For example, with a k-regular interference graph with $k = 20$ and $p = 0.5$, we need around a million nodes for the HT estimate to even have a meaningful value.

However even when HT estimates provide reasonable values, they do not work with dynamic or treatment dependent graphs.

A PROOFS

Lemma A.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 stochastic function f we have*

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

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, $\prod_{i\in S}\left(\frac{z_i}{p}-\frac{1-z_i}{1-p}\right)\prod_{j\in S'} f(z_j)$ 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 $\mathbb{E}\left[f(z_k)\left(\frac{z_k}{p}-\frac{1-z_k}{1-p}\right)\right] = p*\mathbb{E}[f(1)]*\frac{1}{p}+(1-p)*\mathbb{E}[f(0)]*(\frac{-1}{1-p}) = \mathbb{E}[f(1)]-\mathbb{E}[f(0)]$; the RHS when it is non zero simplifies to

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

□

Corollary A.2. *By putting $f(z) = z$ in Lemma A.1we get*

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

Lemma A.3. *Suppose that $\{z_i\}_{i=1..n}$ are mutually independent, with $z_j \sim$ Bernoulli(p). Then, for any subsets S, S',* $\mathbb{E}[\prod_{i\in S} f_i(z_i)\prod_{j\in \S'}\frac{z_j-p}{p}] = \prod_{i\in S\setminus S'}\mathbb{E}[f_i(z_i)]\prod_{k\in S\cap S'}((1-p)(\mathbb{E}[f_k|z_k=1]-\mathbb{E}[f_k|z_k=0]))\,\mathbb{I}[S'\subseteq S]$

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:

$$\begin{aligned}\mathbb{E}[\prod_{i\in S} f_i(z_i)\prod_{j\in \S'}\frac{z_j-p}{p}] &= \mathbb{E}[\prod_{i\in S\setminus S'} f_i(z_i)\prod_{k\in S\cap S'} f_k(z_k)\frac{z_k-p}{p}\prod_{j\in S'\setminus S}\frac{z_j-p}{p}]\\ &= \prod_{i\in S\setminus S'}\mathbb{E}[f_i(z_i)]\prod_{k\in S\cap S'}\mathbb{E}[f_k(z_k)\frac{z_k-p}{p}]\prod_{j\in S'\setminus S}\mathbb{E}[\frac{z_j-p}{p}]\\ &= \prod_{i\in S\setminus S'}\mathbb{E}[f_i(z_i)]\prod_{k\in S\cap S'}((1-p)(\mathbb{E}[f_k|z_k=1]-\mathbb{E}[f_k|z_k=0]))\prod_{j\in S'\setminus S} 0\\ &= \prod_{i\in S\setminus S'}\mathbb{E}[f_i(z_i)]\prod_{k\in S\cap S'}((1-p)(\mathbb{E}[f_k|z_k=1]-\mathbb{E}[f_k|z_k=0]))\,\mathbb{I}[S'\subseteq S]\end{aligned}$$

The exact same argument can be applied to $\mathbb{E}[\prod_{i\in S} z_i\prod_{j\in S'}\frac{p-z_j}{1-p}]$ □

Lemma A.4. *For any sets $S', \mathcal{M}_i$ such that $S'\subseteq \mathcal{N}_i\subseteq \mathcal{M}_i$ and $|S'|\le\beta$ and stochastic functions f_i*

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

Proof.

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

Applying Lemma A.3 with $f_i(z) = z_i$ we get

$$\begin{aligned}
&= \sum_{\substack{S\subseteq\mathcal{M}_i\\|S|\le\beta}} \Bigg[p^{|S'/S|}\prod_{i\in S\setminus S'}\mathbb{E}[f_i(1)](1-p)^{|S'\cap S|}\prod_{i\in S\cap S'}\mathbb{E}[f_i(1)]\mathbb{I}[S\subseteq S'] \\
&\qquad - p^{|S'/S|}\prod_{i\in S\setminus S'}\mathbb{E}[f_i(1)](-p)^{|S'\cap S|}\prod_{i\in S\cap S'}\mathbb{E}[f_i(1)]\mathbb{I}[S\subseteq S']\Bigg] \\
&\overset{(b)}{=} \prod_{i\in S'}\mathbb{E}[f_i(1)]\sum_{\substack{S\subseteq S'\\|S|\le\beta}}\Big[p^{|S'/S|}(1-p)^{|S'\cap S|} - p^{|S'/S|}(-p)^{|S'\cap S|}\Big]
\end{aligned} \tag{S1}$$

(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'

$$\begin{aligned}
&= \prod_{i\in S'}\mathbb{E}[f_i(1)]\sum_{\substack{S\subseteq S'\\|S|\le\beta}} p^{|S'|}\Big[p^{-|S|}(1-p)^{|S|} - p^{-|S|}(-p)^{|S|}\Big] \\
&= \prod_{i\in S'}\mathbb{E}[f_i(1)]\sum_{\substack{S\subseteq S'\\|S|\le\beta}} p^{|S'|}\Big[(\frac{1}{p}-1)^{|S|} - (-1)^{|S|}\Big]
\end{aligned} \tag{S2}$$

If $|S'|\le\beta$, the constraint of $\le\beta$ is redundant. Then by applying binomial theorem we get.

$$= p^{|S'|}\prod_{i\in S'}\mathbb{E}[f_i(1)]\left[\left(1+(\frac{1}{p}-1)\right)^{|S'|} - (1+(-1))^{|S'|}\right] = \prod_{i\in S'}\mathbb{E}[f_i(1)]$$

□

Lemma A.5. *If the set of instrumental variables V is chosen such that $V_j = \frac{Z_j}{p} - \frac{1-Z_j}{1-p}$, then for the pseudo-inverse estimator ($\hat{c}$) in Equation 3, the jth component is given by $\hat{c}_i(j) = Y_i\frac{Z_j}{p} - \frac{1-Z_j}{1-p}$.*

Proof. Note that we are setting $V_j = \frac{Z_j}{p} - \frac{1-Z_j}{1-p}$. Let $X = \mathbb{E}[VZ_{\mathcal{N}_i}^T]$.

Note that $X_{ji} = (\frac{Z_j}{p} - \frac{1-Z_j}{1-p})Z_i$. By Lemma A.1, we know that the the $\mathbb{E}[X_{ji}] = 1\mathbb{I}[j=i]$. Thus the matrix $\mathbb{E}[X]$ is diagonal with 1 for every variable shared between V and $Z_{\mathcal{N}_i}$, and 0 everywhere else. The pseudoinverse of such a matrix is the matrix itself.

The VY_i component of $\hat{c}$ is $(\frac{Z_j}{p} - \frac{1-Z_j}{1-p})Y_i$. Since the pseudo-inverse of X is just diagonal with 1 and 0, with 1 for every variable shared between V and Z; only those components remain. Thus $\hat{c}_i(j) = Y_i\frac{Z_j}{p} - \frac{1-Z_j}{1-p}$ for every index j shared between V and $Z_{\mathcal{N}_i}$. Thus the treatment effect estimate $\tau = \sum\hat{c} = \frac{1}{n}\sum_i Y_i\sum_{j\in V}\left(\frac{z_j}{p} - \frac{(1-z_j)}{(1-p)}\right)$ □

We prove a more general result than the statement in the paper.

Theorem A.1. *Consider a additive model of the form $Y_i(\boldsymbol{z}) = \sum_{S'\subset\mathcal{N}_i} c_{i,S'}\prod_{j\in S'}\mathbb{I}[z_jA_{ij}=1]$. Here each subset of neighbours has an influence which only occurs when all those edges connect to i. Under such a model the GATE effect is given by $\tau = \sum_{S'\subset\mathcal{N}_i} c_{i,S'}\prod_{j\in S'}\mathbb{E}[A_{ij}|\boldsymbol{z}=1]$. If $\mathcal{M}_i\supseteq\mathcal{N}_i$, then $\hat{\tau}^\beta = \frac{1}{n}\sum_i Y_i\sum_{\substack{S\subseteq\mathcal{M}_i\\|S|\le\beta}}\left(\prod_{j\in S}\frac{z_j-p}{p} - \prod_{j\in S}\frac{p-z_j}{1-p}\right)$ is unbiased*

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

$$\begin{aligned}
\mathbb{E}[\hat{\tau}^\beta] &= \mathbb{E}\left[\frac{1}{n}\sum_i Y_i \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \Big(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1-p}\Big)\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_i \sum_{S' \subset \mathcal{N}_i} c_{i,S'} \prod_{j \in S'} \mathbb{I}[z_j A_{ij} = 1] \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \Big(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1-p}\Big)\right] \\
&= \frac{1}{n}\sum_i \mathbb{E}\left[\sum_{S' \subset \mathcal{N}_i} c_{i,S'} \prod_{j \in S'} z_j A_{ij} \sum_{\substack{S \subseteq \mathcal{M}_i \\ |S| \leq \beta}} \Big(\prod_{j \in S} \frac{z_j - p}{p} - \prod_{j \in S} \frac{p - z_j}{1-p}\Big)\right]
\end{aligned}$$

Now applying Lemma A.4 on E1 we get

$$= \frac{1}{n}\sum_i \sum_{S' \subset \mathcal{N}_i} c_{i,S'} \prod_{j \in S'} \mathbb{E}[A_{ij}(1)]\,[1] = \tau(\vec{1}, \vec{0})$$

□

Proof of Proposition 5.2 Unbiasedness of $\hat{\tau}_{OIV}$ follows directly from Theorem A.1 by noting that a) the $\mathcal{M}_i$ in the statement of Proposition 5.2 satisifies the superset criteria in A.1 and b) when $\beta = 1$, $\hat{\tau}^\beta = \hat{\tau}_{OIV}$.

Lemma A.6. *Consider the linear outcome model $Y_i(\boldsymbol{z}) = b_i + c_{ii} Z_i + \sum c_{ij} \mathbb{I}[A_{ij} = 1] Z_j$. For any sets S, consider $Q = Y_i \sum_{j \in S} [\left(\frac{z_j}{p} - \frac{1-z_j}{1-p}\right)]$, we have*

$$\mathbb{E}[Q] = \sum_{j \in S} c_{ij} \mathbb{E}[A_{ij} | Z_j = 1] + \mathbb{I}[i \in S] c_{ii}$$

Proof.

$$\begin{aligned}
\mathbb{E}[Q] &= \mathbb{E}[Y_i \sum_{j \in S} [\left(\frac{z_j}{p} - \frac{1-z_j}{1-p}\right)]] \\
&= \mathbb{E}[\Big(b_i + c_{ii} Z_i + \sum c_{ij} \mathbb{I}[A_{ij} = 1] Z_j\Big) \sum_{j \in S} [\left(\frac{z_j}{p} - \frac{1-z_j}{1-p}\right)]] \\
&= \mathbb{E}[b_i \sum_{j \in S} [\left(\frac{z_j}{p} - \frac{1-z_j}{1-p}\right)]] + \mathbb{E}[c_{ii} Z_i \sum_{j \in S} [\left(\frac{z_j}{p} - \frac{1-z_j}{1-p}\right)]] \\
&\quad + \mathbb{E}[\sum c_{ij} \mathbb{I}[A_{ij} = 1] Z_j \sum_{j \in S} [\left(\frac{z_j}{p} - \frac{1-z_j}{1-p}\right)]]
\end{aligned}$$

Now applying Lemma A.4 we get

$$= 0 + c_{ii} \mathbb{I}[i \subseteq S] + \sum_j c_{ij} \mathbb{E}[A_{ij}(1)] \mathbb{I}[j \subseteq S] = \sum_{j \in S} c_{ij} \mathbb{E}[A_{ij} | Z_j = 1] + \mathbb{I}[i \in S] c_{ii}$$

□

Proof of Proposition 5.6 Applying Lemma A.5, we get that $Y_i \sum_{j \in \mathcal{M}_i^c} [\left(\frac{z_j}{p} - \frac{1-z_j}{1-p}\right)] = \sum_{j \in \mathcal{M}_i^c} c_{ij} \mathbb{E}[A_{ij} | Z_j = 1]$ By the homogeneity assumption, we know that c_{ij} are same. Let it be denoted by k_i. Furthermore by conservation of $\mathcal{M}_i^c$, the edges are always present Thus $Y_i \sum_{j \in \mathcal{M}_i^c} [\left(\frac{z_j}{p} - \frac{1-z_j}{1-p}\right)] = k_i |\mathcal{M}_i^c|$. Next as argued in Section 5.3, to get the treatment effect we can rescale this quantity by $C_i = \frac{1}{p} \sum_j Z_j A_{ij}$ to get unbiased $\hat{\tau}$

A.1 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 these estimates, to handle statistical uncertainty and perform hypothesis tests to verify assumptions. For this purpose, we first argue that these estimators are asymptotically normal.

The generalized central limit theorems [Ross, 2011] assures us that the sum of n bounded random variables R_i, asymptotically behaves like a gaussian distribution if they are mostly independent ; specifically if we construct the dependency graph, then it is not too dense [3]. The dependency graph in out case is provided by the network itself. Hence as long as the underlying interference network is sparse, these estimators are asymptotically normal. The normality of these estimator results suggests a way to do statistical inference. If we can get an upper bound for the variance then we can construct conservative Wald-type intervals [Wasserman, 2006]. We should note however, that since the convergence is asymptotic, the use of the aforementioned variance for confidence intervals is only approximately valid.

Next we provide such conservative bounds for variance of these estimators.

Let the matrix $A \in \{0,1\}^{n\times n}$ denote the dependency graph. We are considering the linear additive model (**A4**). Since we A_{ij} is dependent on Z_j, we can formulate them as $A_{ij} \sim \text{Bernoulli}(q_1)$ if $Z_j = 1$, and $A_{ij} \sim \text{Bernoulli}(q_0)$ if $Z_j = 0$. We assume that the max degree of any node is Δ, . Thus for each node the aforementioned Bernoulli model only applies to Δ nodes. Furthermore we also assume that $|\mathcal{M}_i^c|$ is bounded by $\Delta_{\mathcal{M}^c}$. Finally we have $Z_j \sim \text{Bernoulli}(p)$.

Outcomes Y_i are given by $Y_i = c_i^\top AZ$, where c_i is Δ-sparse (only Δ non-zero entries). We assume that we know an upperbound C for $|c_{ij}|$. We focus on the **UIV Case** as it is more complex and the bound for OIV case can be derived from the bounds in this Section.

The estimator $\hat{\tau}_{UIV}$ is given by:

$$\hat{\tau}_{UIV} = \frac{1}{n}\sum_{i=1}^{n} Y_i \left(\sum_{j\in\mathcal{M}_i^c} \left(\frac{Z_j}{p} - \frac{1-Z_j}{1-p} \right) \sum_{r=1}^{n} Z_r A_{ir} \right).$$

Let $S_i = \sum \left(\frac{Z_j}{p} - \frac{1-Z_j}{1-p} \right)$ and $R = \sum_{r=1}^{n} A_{ir}$. Then:

$$\hat{\tau}_{UIV} = \frac{1}{n}\sum_{i=1}^{n} Y_i S_i R_i.$$

Now

$$\text{Var}(\hat{\tau}_{UIV}) = \frac{1}{n^2}\left[\sum_{i=1}^{n} \text{Var}(Y_i S_i R_i) + 2\sum_{i<j} \text{Cov}(Y_i S_i R_i, Y_j S_j R_j) \right]. \tag{9}$$

First we go about bounding $\text{Var}(Y_i S_i R_i)$. Since $Y_i = \sum_{m\in\mathcal{N}_i} c_{im} \sum_{l=1}^{n} A_{ml} Z_l$ (with $|\mathcal{N}_i| \le \Delta$):

$$|Y_i| \le C \sum_{m\in J_i} \sum_{l=1}^{n} A_{ml} Z_l \le C\Delta.$$

The second moment satisfies:

$$\mathbb{E}[Y_i^2] \le C^2 \mathbb{E}\left[\left(\sum_{l=1}^{n} A_{ml} Z_l \right)^2 \right] \le C^2 \Delta^2 p^2 q_1^2.$$

Next we consider bounding S_i: Each term in S_i is mean 0 and has variance bounded by

$$\text{Var}(S_i) = \sum_i \text{Var}\left(\frac{Z_j}{p} - \frac{1-Z_j}{1-p} \right) \le \Delta_{\mathcal{M}^c} max(\frac{1}{p}, \frac{1}{1-p})$$

[3] For the exact statement we refer the readers to Theorem 3.6 from Ross [2011]

The sum $R_i = \sum_{r=1}^{k} Z_r A_{ir}$ involves Δ terms instead of n. gives:

$$\mathbb{E}[R_i] = \Delta p q_1,$$
$$\text{Var}(R_i) \leq \Delta p q_1 (1 - p q_1).$$

Using Cauchy-Schwarz:

$$\text{Var}(Y_i S_i R_i) \leq \mathbb{E}[(Y_i S_i R_i)^2] \leq C^2 \Delta^3 \Delta_{\mathcal{M}^c} p^5 q_1^4 (1 - p q_1) \frac{1}{\min(p, 1-p)}$$

Next we try bounding Covariance Terms in Equation (9). For $i \neq j$, the covariance $\text{Cov}(Y_i S_i R_i, Y_j S_j R_j)$ is non-zero only if Y_i and Y_j share dependencies. That happens only if there is overlap in $\mathcal{N}_i$ and $\mathcal{N}_j$). Given Δ-sparsity, each Y_i interacts with at most Δ other terms. Thus:

$$\sum_{i<j} \text{Cov}(Y_i S_i R_i, Y_j S_j R_j) \leq n\Delta \cdot \text{Var}(Y_i S_i R_i).$$

Combining terms we get:

$$\text{Var}(\hat{\tau}_{UIV}) \leq \frac{1}{n^2} \left[n \cdot \text{Var}(Y_i S_i R_i) + 2n\Delta \cdot \text{Var}(Y_i S_i R_i) \right].$$

Substituting the bound we get:

$$\text{Var}(\hat{\tau}_{UIV}) \leq \frac{1}{n} (2\Delta + 1) C^2 \Delta^3 \Delta_{\mathcal{M}^c} p^5 q_1^4 (1 - p q_1) \frac{1}{\min(p, 1-p)}$$

We can follow a similar argument for **OIV** case, except in that case $R_i = 1$. Following the same math as before we get the following result

$$\text{Var}(\hat{\tau}_{UIV}) \leq \frac{1}{n} (2\Delta + 1) C^2 \Delta^2 \Delta_{\mathcal{M}}^2 p^4 q_1^3 \frac{1}{\min(p, 1-p)}$$

A.2 MULTI-TRIAL ESTIMATION

By a similar argument as in Section 5.4 (Equation (6)) we can see that under linear additive interference:

$$\mathbb{E}[Y_i] = b_i + c_{ii} p + \sum_j c_{ij} \mathbb{E}[A_{ij}(1)] p$$

which is very similar to the treatment effect, except for the additional term b_i and the scaling by factor of p. We further note that while individual Y_i might be very stochastic and far from their expected value, we can still obtain a good estimate of $E[\sum Y_i]$.

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 their dependency graph is not too dense, then the variance normalized sum approaches a normal distribution. If we consider Y_i to be these random variables, their dependency graph is represented by the matrix $\boldsymbol{A}$. If $\boldsymbol{A}$ is not too dense under any counterfactual, then $\frac{1}{n} \sum_i^n Y_i$ is asymptotically normal with mean $\frac{1}{n} \sum_i^n \mathbb{E}[Y_i]$.

On the other hand we know that $\mathbb{E}[Y_i]$ is linear in p. Let $F(Y) = \frac{1}{n} \sum_i Y$, then $\mathbb{E}[F(Y)] = \frac{1}{n} \sum_i [c_{ii} + \sum_j c_{ij} \mathbb{E}[A_{ij}(1)] p] = \frac{1}{n} (\sum c_{ii}) + \tau p$

Remark A.7. This holds true for more complex interaction models. More specifically if the set of all possible neighbours under all possible interactions (i.e. $\mathcal{M}_i^c$))is bounded by a number β, then $\frac{1}{n} \sum_i^n Y_i$ asymptotically converges to a polynomial of order β in p.

Under the linear interference model, we can conduct two experiments at two different randomization probabilities p_1 and p_2, and fit a linear function in p. Let that function be $\hat{F}$. By the earlier argument $\hat{F}(p)$ is unbiased and consistent estimate of $F(p)$. By definition, global treatment effect τ is given by $(F(1) - F(0))$. Thus we have the following estimator

$$\hat{\tau}_{MULTI} = \hat{F}(1) - \hat{F}(0)$$

Proposition A.8. *Under assumptions **A1-4**, and assuming multiple independent trials, $\hat{\tau}_{MULTI}$ is an unbiased estimate of the treatment effect τ*

B TREATMENT DEPENDENT NETWORKS

B.1 FAILURE OF HT ESTIMATION

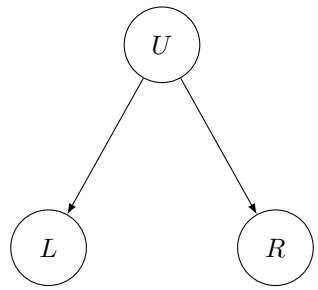


Figure 5

edge UL exists if and only if $Z_U = 1$ otherwise the edge UR will exist. However outcomes at L and R, i.e. (Y_L, Y_R) respectively are independent of treatment at U and only depend on treatment at self with the effect being constant α i.e. the outcomes are $Y_{L/R}(1) = Y_{L/R}(0) + \alpha_{L/R}$. All treatments are perfectly randomized with probability $q = 0.5$.

We consider the TTE (GATE) between $Z = \vec{0}$ and $Z = \vec{1}$ with the HT estimate here. By symmetry, we can consider only U, L with the U, R case analogous. We have 4 potential treatments with corresponding values for the HT estimator being

$$\text{For } Z_U = 1, Z_L = 1 \text{ we have} \qquad Y_L(1)(\frac{1}{0.25})$$

$$\text{For } Z_U = 1, Z_L = 0 \text{ we have} \quad Y_L(0)(\frac{1*0}{0.25} - \frac{0*1}{0.25}) = 0$$

$$\text{For } Z_U = 0, Z_L = 1 \text{ we have} \qquad Y_L(1)(\frac{1}{0.5})$$

$$\text{For } Z_U = 0, Z_L = 0 \text{ we have} \qquad Y_L(0)(-\frac{1}{0.5})$$

The expected value for the contribution of node L in all possibilities is $(Y_L(1) + \frac{Y_L(1)-Y_L(0)}{2})$.

Similarly, the contribution from node R is $(-Y_R(0) + \frac{Y_R(1)-Y_R(0)}{2})$.

Hence, the expected value of the estimator is given by $\frac{Y_L(1)-Y_R(0)}{2} + \frac{\alpha_L+\alpha_R}{4}$.

B.2

We explain here a bit more formally the issue with HT estimation. For simplicity we will consider the linear interference model. First we would like to begin with the following result from Sussman and Airoldi [2017].

Under assumption **A1-2,A4**, the expected value of the HT estimate and the HATE estimator $\hat{\tau}_{HATE} = \frac{1}{n}\sum_i Y_i \sum_{j\in\mathcal{N}_i}\left(\frac{z_j}{p} - \frac{(1-z_j)}{(1-p)}\right)$ is the same.

We first have a look at how assumption **A4** affects τ_{HT}. Substituting **A4** in Equation (2), τ_{HT} can be expressed as

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

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 Lemma A.1). Therefore, τ_{HT} can be simplified as

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

which is a linear combination of in the terms z_j/p and $(1-z)/(1-p)$; however this expression cannot be computed from only the graph and observed outcomes Y_i.

We will rewrite this 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 $\mathcal{N}_i$, we get

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

which motivates the following estimator:

$$\hat{\tau} = \frac{1}{n}\sum_i Y_i \sum_{j\in\mathcal{M}_i}\left(\frac{z_j}{p} - \frac{(1-z_j)}{(1-p)}\right). \tag{10}$$

Now comes the crucial issue: the above derivation obscures the fact that $\mathcal{N}_i$ depended on $\boldsymbol{z}$. Specifically a node $j \in \mathcal{N}_i$ if $A_{ij} = 1$. But since A_{ij} are themselves potential outcome functions dependent on $\boldsymbol{z}$, this neighbourhood is not static. One includes node j in the above sum if the edge was observed, and the probability of observing the edge is different for $z_j = 0$ and $z_j = 1$. Hence the more appropriate expression is

$$\hat{\tau} = \frac{1}{n}\sum_i Y_i \sum_j \left(\mathbb{I}[A_{ij} = 1|z_j = 1]\left(\frac{z_j}{p}\right) - \mathbb{I}[A_{ij} = 1|z_j = 0]\left(\frac{(1-z_j)}{(1-p)}\right)\right). \tag{11}$$

Unlike $\left(\frac{z_j}{p} - \frac{1-z}{1-p}\right)$ this term is not necessarily mean 0, but is instead $\mathbb{E}[A_{ij}|z_j = 1] - \mathbb{E}[A_{ij}|z_j = 0]$, and including this term in the regression ends up biasing the estimate. This is why we needed either $\mathcal{M}_i$ (superset) or $\mathcal{M}_i^c$ (subset). In either case the neighbourhood over which the $\left(\frac{z_j}{p} - \frac{1-z}{1-p}\right)$ terms are added is not dependent on $\boldsymbol{z}$. $\hat{\tau}_{OIV}$ includes edges irrespective of their A_{ij} value (or more specifically skips terms only if $A_{ij}(z_j) = 0$ always). On the other hand $\hat{\tau}_{UIV}$ only includes edges if $A_{ij}(z_j) = 1$ always (and hence independent of $\boldsymbol{z}$).

C EXPERIMENTAL DETAILS

C.1 SYNTHETIC GRAPHS

We sample different random Graphs and run repeated experiments on these graphs with randomized bernoulli treatment assignment. The baselines include the POLY(Prop/Num) estimator is a polynomial regression on the exposure as computed by the fraction/number of treated nodes in the neighbourhood. The DM estimator signifies the classic difference in mean/ SUTVA estimator which is is simply the average outcomes on treated vs un-treated units. The ER graphs are made with an expected neighbourhood of size 20. The outcome model is similar to the potential outcomes model as in Cortez et al. [2022]:

$$Y_i(\boldsymbol{z}) = c_{i,\emptyset} + \sum_{j\in\mathcal{N}_i}\tilde{c}_{i,1} z_j + \sum_{\ell=2}^{\beta}\left(\frac{\sum_{j\in\mathcal{N}_i}\tilde{c}_{ij,2} a_{ij} z_j}{\sum_{j\in\mathcal{N}_i}\tilde{c}_{ij,2}}\right)^{\ell}, \tag{12}$$

where $i \neq j$, $\tilde{c}_{ij,2} = v_{i,2}|\mathcal{N}_i|/\sum_{k:(k,j)\in E}|\mathcal{N}_k|$. The coefficient $c_{i,\emptyset}, \tilde{c}_{i,1}, v_{i,2}$ are obtained as random variables.