Exploiting Neighborhood Interference with Low Order Interactions under Unit Randomized Design

Mayleen Cortez-Rodriguez
Center for Applied Math
Cornell University
Ithaca, NY 14850
mec383@cornell.edu

Matthew Eichhorn
Center for Applied Math
Cornell University
Ithaca, NY 14850
meichhorn@cornell.edu

Christina Lee Yu
Dept of Operations Research
Cornell University
Ithaca, NY 14850
cleeyu@cornell.edu

Abstract

Network interference, where the outcome of an individual is affected by the treatment of others in their social network, is pervasive in real-world settings. However, it poses a challenge to estimating causal effects. We consider the task of estimating the total treatment effect (TTE), or the difference between the average outcomes of the population when everyone is treated versus when no one is, under network interference. Under a non-uniform Bernoulli randomized design, we utilize knowledge of the network structure to provide an unbiased estimator for the TTE when network interference effects are constrained to low-order interactions among neighbors of an individual. We make no assumptions on the graph other than bounded degree, allowing for well-connected networks that may not be easily clustered. We derive a bound on the variance of our estimator and show in simulated experiments that it performs well compared with standard TTE estimators.

1 Introduction

Accurately estimating causal effects is relevant in numerous applications, from pharmaceutical companies researching the efficacy of a new medication, to policy makers understanding the impact of social welfare programs, to social media companies evaluating the impact of different recommendation algorithms on user engagement across their platforms. The *total treatment effect* (TTE), the difference between the average individual's outcomes when everyone is treated versus when no one is treated, is a useful estimand in these settings, as it provides an approximate "return on investment" from a large-scale rollout of the treatment. To arrive at such an estimate, a company or agency may design an experiment where they randomly assign subsets of the population to treatment (e.g. new medication) and to control (e.g. a placebo) and draw conclusions based on the observed outcomes of the participants (e.g. health outcomes).

The techniques and guarantees for estimating causal effects in classical causal inference heavily rely upon the stable unit treatment value assumption (SUTVA), which posits that the outcome of each individual is independent of the treatment assignment of all other individuals (12). Unfortunately, SUTVA is violated in all of the above applications due to *network interference*: one's outcome is impacted not only by their treatment, but also by the treatment of their peers. Distinguishing between the direct effect of treatment on an individual and the network effect of others' treatment can be challenging. This has resulted in a growing literature on causal inference in the presence of such *interference* or *spillover effects*.

We consider the task of estimating the TTE under the presence of network interference. In particular, our work assumes *neighborhood interference*, under which an individual is affected by the treatment of their direct neighbors but is unaffected by the treatment of those individuals outside their neighborhood. Furthermore, we focus on an experimental setting under *unit randomized designs*,

NeurIPS 2022Workshop on Causal Machine Learning for Real-World Impact (CML4Impact 2022).

wherein individuals are independently assigned to either the treatment or control group. This is in contrast to cluster randomized designs, which have been proposed as an approach to address network interference for randomized experimental design but which may not be feasible in practice due to an incompatibility with existing experimental platforms or to policy regulations.

Related Work. The literature has largely taken two approaches towards estimating the total treatment effect under network interference. The majority of work in the non-parametric approach assumes *partial interference*: the population can be partitioned into groups, and network interference occurs within but not across groups (1; 3; 9; 10; 13; 16; 20). Given knowledge of the groups, one typically randomly assign groups to different treatment saturation levels, for example jointly assigning an entire group to either treatment or control. Then, classical approaches such as a Difference-in-Means or Horvitz-Thompson estimator can be used to estimate the TTE. The asymptotic consistency of these estimators relies on the number of groups going to infinity. In practice, even networks with an obvious clustering may not have a sufficiently large number of groups to result in useful estimates. When partial interference is not satisfied, standard estimators incur bias that scales with the number of edges between groups. Therefore, the goal of cluster-based randomized designs in this setting is to leverage inherent structures to find a clustering that minimizes the number of edges between clusters (8; 7; 18; 19). Finding such a clustering can be computationally expensive, difficult to implement in existing experimentation platforms, or unfair due to nonuniform treatment exposures. For this reason, we limit our attention to unit randomized design.

The other common approach is to impose strong structural properties on the potential outcomes model. The most common assumption is that the potential outcomes are linear with respect to a particular statistic of the treatment vector (2; 4; 5; 8; 11; 17). This approach reduces the number of unknown parameters in the potential outcomes function so that causal estimation can be recast as a linear regression problem. This lends itself naturally to using a least squares estimate, turning the attention to minimizing the variance of the estimate by using an appropriate randomized design. A drawback of this approach is that it frequently assumes anonymous interference, which imposes homogeneity amongst the network effects. Our work removes need for linearity by providing a simple, unbiased estimator for the TTE under a potential outcomes model that is polynomial with respect to the treatment vector. Moreover, our model allows for heterogeneity in the influence of different sets of treated neighbors. This removes the restrictive anonymous interference assumption, so strictly generalizes beyond the typical parametric model classes.

Some of the most similar work to ours is found in (21) and (6). The former provides an estimator for the TTE with neighborhood interference under a heterogeneous linear model, also called the assumptions of additivity of main effects and interference in (14). Interestingly, their method requires no knowledge of the underlying network and instead utilizes measurements over two time steps. In (6), the authors generalize this work beyond linear to polynomial potential outcomes models, requiring measurements at multiple time-steps. Our work considers the same potential outcomes model as them, but in scenarios where we have knowledge of the network. In particular, we assume knowledge of all the neighborhood sets. Our results apply to settings where we only have access to measurements at a single time-step, or where data is observational, both of which are not addressed in (6) or (21).

2 Model

Causal Network. Let $[n] := \{1, \ldots, n\}$ denote the underlying population of n individuals. We model the network effects in the population as a directed graph over the individuals with edge set $E \subseteq [n] \times [n]$. An edge $(j, i) \in E$ signifies that the treatment assignment of individual j affects the outcome of individual i. As an individual's own treatment is likely to affect their outcome, we expect self-loops in this graph. We use $\mathcal{N}_i := \{j \in [n] : (j, i) \in E\}$ to denote the in-neighborhood of an individual i. Note that this definition allows $i \in \mathcal{N}_i$. Our variance bound is parameterized by the network degree. We let d_{in} denote the maximum in-degree of any individual and d_{out} denote the maximum out-degree.

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

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

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

$$Y_i(\mathbf{z}) = \textstyle\sum_{\mathcal{S}' \subseteq [n]} c_{i,\mathcal{S}'} \prod_{j' \in \mathcal{S}'} z_{j'}, \tag{1}$$

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

So far, the potential outcomes model described in (1) is completely general. However, it is parameterized by 2^n coefficients $\{c_{i,\mathcal{S}'}\}$, which makes it untenable in most settings. To combat this, we impose some structural assumptions on these coefficients. First, we observe that the populations of interest can be quite large (e.g. the population of an entire country), and their influence networks may have high diameter. Throughout most of the paper, we assume that individuals' outcomes are influenced only by their immediate in-neighborhood.

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

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

Assumption 2 (Low Polynomial Degree)**.** *Each potential outcome function $Y_i(\mathbf{z})$ has degree at most β. In our notation, $c_{i,\mathcal{S}'} = 0$ whenever $|\mathcal{S}'| > \beta$.*

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

Under Assumptions 1 and 2, we may re-express $Y_i(\mathbf{z})$ from (1) in the form,

$$Y_i(\mathbf{z}) = \sum_{\substack{\mathcal{S}' \subseteq \mathcal{N}_i \\ |\mathcal{S}'| \le \beta}} c_{i,\mathcal{S}'} \prod_{j \in \mathcal{S}'} z_j. \tag{2}$$

The number of unknown parameters in this model is $\sum_{i \in [n]} \sum_{k=0}^{\beta} \binom{|\mathcal{N}_i|}{k}$, which scales as $n d_{\text{in}}^{\beta}$. As noted above, taking $\beta = 1$ corresponds to the heterogeneous linear outcomes model in (21). This low degree assumption will not generally admit threshold models or saturation models, both of which would require the degree of $Y_i(\mathbf{z})$ to be $|\mathcal{N}_i|$.

Our variance bounds utilizes an upper bound on the treatment effects for each individual. We define $Y_{\max}$ such that

$$Y_{\max} := \max_{i \in [n]} \textstyle\sum_{\substack{\mathcal{S} \subseteq \mathcal{N}_i \\ |\mathcal{S}| \le \beta}} |c_{i,\mathcal{S}}|.$$

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

Assumption 3 (Observation Noise)**.** *Assume that the observations of individual's outcomes are perturbed by independent Gaussian noise such that $Y_i^{obs} = Y_i(\mathbf{z}) + \epsilon_i$ for $\epsilon_i \sim N(0, \sigma^2)$.*

Causal Estimand and Randomized Design. Throughout this paper, we concern ourselves with estimating the total treatment effect (TTE). This quantifies the difference between the average of individual's outcomes when the entire population is treated versus the average of individual's outcomes when the entire population is untreated:

$$\text{TTE} := \frac{1}{n}\sum_{i=1}^{n}\big(Y_i(\mathbf{1}) - Y_i(\mathbf{0})\big), \tag{3}$$

where $\mathbf{1}$ represents the all 1s vector and $\mathbf{0}$ represents the zero vector. Plugging in our parameterization from equation (2), we may re-express the total treatment effect as

$$\text{TTE} = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_i \\ 1\le|\mathcal{S}'|\le\beta}} c_{i,\mathcal{S}'}. \tag{4}$$

Since exposing individuals to treatment can have a deleterious and irreversible effect on their outcomes, we wish to estimate the total treatment effect after treating a small random subset of the population. Throughout the paper, we focus on a *non-uniform* Bernoulli design, wherein each individual i is independently assigned treatment with probability $p_i \in [p, 1-p]$ for $p \in (0, 0.5]$, i.e. each $z_i \sim \text{Bern}(p_i)$. This randomized design is straightforward to implement and understand.

3 Estimator

In this section we introduce the estimator that forms the basis for much of our work. We consider an experimental setting utilizing a non-uniform Bernoulli design; each individual i is treated independently with probability p_i. Then, our estimator is given by

$$\widehat{\text{TTE}} = \frac{1}{n}\sum_{i=1}^{n} Y_i^{\text{obs}}(\mathbf{z}) \sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i \\ |\mathcal{S}|\le\beta}} g(\mathcal{S}) \prod_{j\in\mathcal{S}}\left(\frac{z_j}{p_j} - \frac{1-z_j}{1-p_j}\right), \tag{5}$$

where we define $g : 2^{[n]} \to \mathbf{R}$ such that $g(\mathcal{S}) = \prod_{s\in\mathcal{S}}(1-p_s) - \prod_{s\in\mathcal{S}}(-p_s)$ for each $\mathcal{S} \subseteq [n]$. Note that this estimator can be evaluated in $O(nd_{\text{in}}^{\beta})$ time and only utilizes structural information about the graph (not any influence coefficients $c_{i,\mathcal{S}}$). We remark that this estimator is a special case of the *pseudoinverse estimator* first introduced by Swaminathan et. al. (15).

Structurally, the estimator takes the form of a weighted average of the outcomes $Y_i(\mathbf{z})$ of each individual i, where the weights themselves are functions of the treatment assignments of all members j of the in-neighborhood $\mathcal{N}_i$. To make use of the low-order interference assumption, the estimator separately scales the effect of treatment of each sufficiently small subset of $\mathcal{N}_i$ using the scaling function $g(\mathcal{S})$. The definition of this $g(\mathcal{S})$ ensures the unbiasedness of the estimator and recovers the Horvitz-Thompson estimator when $\beta \ge d_{\text{in}}$. In the special case of a uniform treatment probability $p_i = \hat{p}$ across all nodes, we can simplify this estimator to show that it is only a function of the number of treated individuals in i's neighborhood, and not their specific identities.

The following theorem summarizes the key properties of our estimator.

Theorem 1. *Under a potential outcomes model satisfying the neighborhood interference assumption with polynomial degree at most β, the estimator defined in* (5) *is unbiased with variance bounded by*

$$\frac{d_{\text{in}}\, d_{\text{out}}\, {Y_{\max}}^2}{n} \cdot \left(\frac{d_{\text{in}}^2}{p(1-p)}\right)^{\beta} + \frac{\sigma^2}{n}\left(\frac{d_{\text{in}}}{p(1-p)}\right)^{\beta},$$

where each $p_i \in [p, 1-p]$ and $p \in (0, 0.5]$.

Notably, a sequence of networks with $n \to \infty$ and $d = o(\log n)$ has variance asymptotically approaching 0. We defer the proof of this theorem to Appendix A.

Discussion. Our estimator is a linear weighted estimator which takes the form of $\frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z})w_i(\mathbf{z})$ for a specially constructed weight function $w_i(\mathbf{z})$. We can contrast our estimator to the commonly used Horvitz-Thompson estimator, which takes the form of

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^{\text{obs}}(\mathbf{z})\left(\frac{\mathbb{I}(\mathbf{z}\text{ treats all of }\mathcal{N}_i)}{\Pr(\mathbf{z}\text{ treats all of }\mathcal{N}_i)} - \frac{\mathbb{I}(\mathbf{z}\text{ does not treat all of }\mathcal{N}_i)}{\Pr(\mathbf{z}\text{ does not treat all of }\mathcal{N}_i)}\right).$$

In the special case when $|\mathcal{N}_i| \le \beta$ the restriction $|\mathcal{S}| \le \beta$ is always satisfied for every $\mathcal{S} \subseteq \mathcal{N}_i$, such that by an application of the Binomial theorem, it follows that $w_i(\mathbf{z})$ reduces to

$$w_i(\mathbf{z}) = \sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i \\ |\mathcal{S}|\le\beta}} g(\mathcal{S}) \prod_{j\in\mathcal{S}} \left(\frac{z_j}{p_j} - \frac{1-z_j}{1-p_j}\right) = \prod_{j\in\mathcal{N}_i} \frac{z_j}{p_j} - \prod_{j\in\mathcal{N}_i} \frac{1-z_j}{1-p_j}$$

which is equivalent to the weight used in the Horvitz-Thompson estimator under Bernoulli randomized design. As a result, when β is sufficiently large relative to the degree of the nodes in the graph, our estimator is very similar to the Horvitz-Thompson estimator, only differing for the nodes which have graph degree larger than β. In this sense, under Bernoulli randomization, our estimator can be viewed as a generalization of Horvitz-Thompson to account for low polynomial degree structure, which is most relevant for simplifying the potential outcomes associated to high degree vertices.

We can compare against the variance of Horvitz-Thompson under a Bernoulli design. In the simple setting of a d-regular graph and uniform Bernoulli(p) randomization, (19) showed that the Horvitz-Thompson estimator has a variance that is lower bounded by $\Omega(1/np^d)$. In contrast, the variance of our estimator only scales polynomially in the degree d, but exponentially in the polynomial degree β, which is achieved by simply changing the estimator, without requiring any additional clustering structure of the graph and without utilizing complex randomized designs. This is a significant gain when the polynomial degree β is significantly lower than the graph degree d. The simplest setting of $\beta = 1$ already expresses all potential outcomes models which satisfy additivity of main effects and additivity of interference, as defined in (14); this subsumes all linear models which are commonly used in the practical literature, yet which require additional homogeneity assumptions.

4 Experimental Results

Using computational experiments on simulated data, we compare the performance of our estimator with existing estimators. Using an Erdős-Rényi model, we generate random directed graphs of n nodes for a population of n individuals. Figure 1 shows results from networks made using the Erdős-Rényi model with n nodes and probability $p_{\text{edge}} = 10/n$ of an edge existing between any two nodes. Hence, the expected in-degree and out-degree of each node is 10. For degree β, we construct the same potential outcomes model as in (6):

$$Y_i(\mathbf{z}) = c_{i,\emptyset} + \sum_{j\in\mathcal{N}_i} \tilde{c}_{ij} z_j + \sum_{\ell=2}^{\beta} \left(\frac{\sum_{j\in\mathcal{N}_i} \tilde{c}_{ij} z_j}{\sum_{j\in\mathcal{N}_i} \tilde{c}_{ij}}\right)^{\ell}, \tag{6}$$

where $c_{i,\emptyset} \sim U[0,1]$, $\tilde{c}_{ii} \sim U[0,1]$, and for $i \neq j$, $\tilde{c}_{ij} = v_j |\mathcal{N}_i| / \sum_{k:(k,j)\in E} |\mathcal{N}_k|$ for $v_j \sim U[0,r]$, where r denotes a hyperparameter that governs the magnitude of the network effects relative to the direct effects. We represent the magnitude of individual j's influence by the parameter v_j. This influence is shared among individual j's out-neighbors proportional to their in-degrees. For simplicity, we assume no observation noise, i.e. $\sigma = 0$.

Other Estimators. We compare the performance of our estimator with the performance of least-squares regression and difference-in-means estimators. Let U_i denote the number of individuals in $\mathcal{N}_i \setminus \{i\}$ assigned to treatment, and let $\tilde{U}_i$ denote the number of neighbors individuals in $\mathcal{N}_i \setminus \{i\}$ assigned to control. For some user-defined tolerance $\lambda \in [0,1]$, the difference-in-means estimator is given by

$$\widehat{\text{TTE}}_{\text{DM}(\lambda)} = \frac{\sum_{i\in[n]} z_i \mathbb{I}(U_i \ge \lambda) Y_i(\mathbf{z})}{\sum_{i\in[n]} z_i \mathbb{I}(U_i \ge \lambda)} - \frac{\sum_{i\in[n]} (1-z_i) \mathbb{I}(\tilde{U}_i \ge \lambda) Y_i(\mathbf{z})}{\sum_{i\in[n]} (1-z_i) \mathbb{I}(\tilde{U}_i \ge \lambda)}. \tag{7}$$

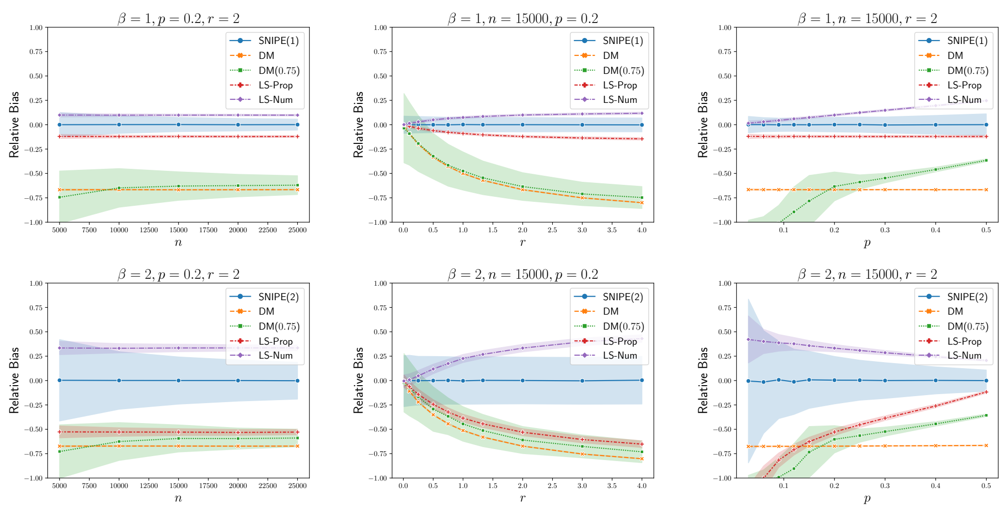
We set $\lambda = 1$ and $\lambda = 0.75$ for our experiments. Note that $\widehat{\text{TTE}}_{\text{DM}(\lambda)}$ counts an individual i's outcome only when at least λ of their neighborhood is assigned to the same treatment as them. We will denote $\widehat{\text{TTE}}_{\text{DM}} = \widehat{\text{TTE}}_{\text{DM}(1)}$, as it corresponds to the classical difference in means estimator.

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

$$Y_i(\mathbf{z}) = g(z_i, \bar{z}_i) = \left(\rho + \sum_{k=1}^{\beta} \gamma_k X_i^k\right) + z_i\left(\tilde{\rho} + \sum_{k=1}^{\beta-1} \tilde{\gamma}_k X_i^k\right), \tag{8}$$

for some covariate X_i. We consider two variations. In the first, we set X_i equal to the number of treated neighbors. In the second, we let X_i equal the proportion of treated neighbors. In both cases, we do not include i in its neighborhood. The two sets of coefficients $(\rho, \gamma_1, \ldots \gamma_\beta)$ and $(\tilde{\rho}, \tilde{\gamma}_1, \ldots \tilde{\gamma}_\beta)$ allow for the model to be different when i is treated vs not treated, and since we only allow up to degree β interactions, the second summation stops at $\beta - 1$. Overall, there are $2\beta + 1$ coefficients in the model. Using least-squares regression, we determine the set of coefficients minimizing the least-squares predictive error on the data set $\{z_i, X_i, Y_i(\mathbf{z})\}_{i \in [n]}$. These coefficients define an estimate for the function $\hat{g}$ in Equation 8. Given the fitted estimate $\hat{g}$, the estimate for the total treatment effect, denoted by $\widehat{\text{TTE}}_{\text{LS-Num}}$ and $\widehat{\text{TTE}}_{\text{LS-Prop}}$ for the two linear models respectively, is computed by simply substituting in the fitted function $\hat{g}$ into the definition of the TTE.

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



(a) Varying population size (b) Varying direct:indirect effects (c) Varying treatment budget

Figure 1: Plots visualizing the performance of various TTE estimators under Bernoulli design on Erdős-Rényi networks for both linear and quadratic potential outcomes models. The height of each line on a plot depicts the experimental relative bias of the estimator and the shaded width depicts the experimental standard deviation.

Figure 1 visualizes the effects of various network or estimator parameters on the performance of each of the four benchmark TTE estimators and our estimator $\widehat{\text{TTE}}_{\text{SNIPE}(\beta)}$, all under Bernoulli randomized design. In particular, we consider the effects of the population size (n), the treatment budget (p), the ratio between the network and direct effects (r), and the degree of the potential outcomes model (β). Figure 1 shows the bias and empirical standard deviation of each estimator, where the values are all normalized by the magnitude of the true TTE.

The top row of plots in Figure 1 features results for a linear ($\beta = 1$) potential outcomes model while the bottom row shows results for a quadratic ($\beta = 2$) potential outcomes model. As expected, our estimator, shown in blue, has no relative bias and its variance decreases as n increases. With the exception of the modified difference-in-means estimator $\widehat{TTE}_{\text{DM}(0.75)}$ in green, the variances of the other estimators are lower than ours. However, the biases of the other estimators are larger than the

[1]We ran all experiments on a Linux-based machine with 20 CPU(s) and 10 cores. The Python scripts for the experiments and the data used in our results are available at: `https://github.com/mayscortez/low-order-unitRD`

variance of our unbiased estimator overall. Moreover, as r increases, the networks effects are more significant than the direct effects and we see the biases of the other estimators grow larger. Note that the variance of our estimator remains relatively constant as r varies. When r is close to 0, there are essentially no network effects, SUTVA holds and as expected, all the estimators are unbiased.

Acknowledgments and Disclosure of Funding

We gratefully acknowledge financial support from the National Science Foundation grants CCF-1948256 and CNS-1955997 and the National Science Foundation Graduate Research Fellowship Program grant DGE-1650441. Dr. Yu is also supported by an Intel Rising Stars award and a JPMorgan Faculty Research award. Mayleen is also supported in part by the Sloan Foundation grant 90855.

References

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

[2] 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.

[3] 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. PMLR, 22–25 Jul 2020. URL `https://proceedings.mlr.press/v115/bhattacharya20a.html`.

[4] 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.

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

[6] Mayleen Cortez, Matthew Eichhorn, and Christina Lee Yu. Graph agnostic estimators with staggered rollout designs under network interference. *arXiv preprint arXiv:2205.14552*, 2022. doi: 10.48550/ARXIV.2205.14552. URL `https://arxiv.org/abs/2205.14552`.

[7] 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.

[8] 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. International World Wide Web Conferences Steering Committee, 2015.

[9] Michael G. Hudgens and M. Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103:832–842, 2008. URL `https://EconPapers.repec.org/RePEc:bes:jnlasa:v:103:y:2008:m:june:p:832-842`.

[10] 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.

[11] Ben M. Parker, Steven G. Gilmour, and John Schormans. 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, 2017. doi: https://doi.org/10.1111/rssc.12170. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssc.12170`.

[12] Donald B. Rubin. Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, 75(371):591–593, 1980. ISSN 01621459. URL `http://www.jstor.org/stable/2287653`.

[13] Michael E Sobel. What do randomized studies of housing mobility demonstrate? *Journal of the American Statistical Association*, 101(476):1398–1407, 2006. doi: 10.1198/016214506000000636. URL https://doi.org/10.1198/016214506000000636.

[14] 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.

[15] 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.

[16] 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.

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

[18] Johan Ugander and Hao Yin. Randomized graph cluster randomization. *arXiv preprint arXiv:2009.02297*, 2020. doi: 10.48550/ARXIV.2009.02297. URL https://arxiv.org/abs/2009.02297.

[19] Johan Ugander, Brian Karrer, Lars Backstrom, and Jon Kleinberg. 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, 2013.

[20] 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.

[21] Christina Lee Yu, Edoardo Airoldi, Christian Borgs, and Jennifer Chayes. Estimating total treatment effect in randomized experiments with unknown network structure. *arXiv preprint arXiv:2205.12803*, 2022.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] See discussion after assumption 2 and last paragraph of section 3
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] This is implicit in the introduction.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes] Appendix A
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] URL provided in footnote
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects [N/A]

A Proof of Theorem 1

The key insight that we use in the construction of this unbiased estimator comes from the following lemma.

Lemma 1. *Suppose that $\{z_j\}_{j\in[n]}$ are mutually independent, with $z_j \sim$ Bernoulli(p_j). Then, for any $\mathcal{S}, \mathcal{S}' \subseteq [n]$,*

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

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

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

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

The critical feature of this lemma, as will become apparent in the subsequent proofs, is that this indicator function simplifies sums over arbitrary sets to sums over subsets $\mathcal{S}\subseteq\mathcal{S}'$. This additional structure permits simplification through techniques including the binomial theorem and Möbius inversion.

Applying linearity of expectation, we have

$$\mathbb{E}\Big[\widehat{\text{TTE}}\Big] = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_i\\|\mathcal{S}'|\le\beta}} c_{i,\mathcal{S}'}\sum_{\substack{\mathcal{S}\in\mathcal{N}_i\\|\mathcal{S}|\le\beta}} g(\mathcal{S})\cdot\mathbb{E}\Big[\prod_{j\in\mathcal{S}}\Big(\frac{z_j}{p_j}-\frac{1-z_j}{1-p_j}\Big)\prod_{j'\in\mathcal{S}'} z_{j'}\Big].$$

Applying Lemma 1, this simplifies to

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_i\\1\le|\mathcal{S}'|\le\beta}} c_{i,\mathcal{S}'}\sum_{\mathcal{S}\subseteq\mathcal{S}'} g(\mathcal{S})\prod_{j'\in\mathcal{S}'\setminus\mathcal{S}} p_{j'}.$$

Next, substituting in the definition of $g(\mathcal{S})$, and rearranging we have

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_i\\1\le|\mathcal{S}'|\le\beta}} c_{i,\mathcal{S}'}\prod_{j'\in\mathcal{S}'} p_{j'}\sum_{\mathcal{S}\subseteq\mathcal{S}'}\Big(\prod_{j\in\mathcal{S}}\frac{1-p_j}{p_j} - (-1)^{|\mathcal{S}|}\Big).$$

Applying the binomial theorem allows us to cancel out the $(-1)^{|\mathcal{S}|}$ terms. For the remaining terms we may rewrite

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_i\\1\le|\mathcal{S}'|\le\beta}} c_{i,\mathcal{S}'}\prod_{j'\in\mathcal{S}'} p_{j'}\sum_{\mathcal{S}\subseteq\mathcal{S}'}\prod_{j\in\mathcal{S}}\Big(\frac{1}{p_j}-1\Big)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_i\\1\le|\mathcal{S}'|\le\beta}} c_{i,\mathcal{S}'}\prod_{j'\in\mathcal{S}'} p_{j'}\sum_{\mathcal{S}\subseteq\mathcal{S}'}\sum_{\mathcal{T}\subseteq\mathcal{S}}(-1)^{|\mathcal{S}|-|\mathcal{T}|}\cdot\prod_{j\in\mathcal{T}}\frac{1}{p_j}.$$

Finally, applying Möbius inversion on the boolean poset $(\mathcal{P}(\mathcal{S}'),\subseteq)$, we have

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_i\\1\le|\mathcal{S}'|\le\beta}} c_{i,\mathcal{S}'}\prod_{j'\in\mathcal{S}'} p_{j'}\cdot\prod_{j\in\mathcal{S}'}\frac{1}{p_j} \qquad = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_i\\1\le|\mathcal{S}'|\le\beta}} c_{i,\mathcal{S}'} \qquad = \text{TTE}.$$

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

Lemma 2. *For any* $\mathcal{S} \subseteq [n]$, $|g(\mathcal{S})| \leq 1$.

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

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

□

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

Lemma 3. *Suppose that* $\{z_j\}_{j\in[n]}$ *are mutually independent, with* $z_j \sim \text{Bernoulli}(p_j)$. *Then, for any* $\mathcal{S}, \mathcal{S}', \mathcal{T}, \mathcal{T}' \subseteq [n]$,

$$\bigg|\text{Cov}\Big[\prod_{j\in\mathcal{S}}\Big(\frac{z_j}{p_j} - \frac{1-z_j}{1-p_j}\Big)\prod_{j'\in\mathcal{S}'}z_{j'}\ ,\ \prod_{k\in\mathcal{T}}\Big(\frac{z_k}{p_k} - \frac{1-z_k}{1-p_k}\Big)\prod_{k'\in\mathcal{T}'}z_{k'}\Big]\bigg| \leq (p(1-p))^{-\beta}.$$

Proof. We reason separately about the two terms in the covariance expansion. By Lemma 1,

$$\mathbb{E}\Big[\prod_{j\in\mathcal{S}}\Big(\frac{z_j}{p_j} - \frac{1-z_j}{1-p_j}\Big)\prod_{j'\in\mathcal{S}'}z_{j'}\Big]\mathbb{E}\Big[\prod_{k\in\mathcal{T}}\Big(\frac{z_k}{p_k} - \frac{1-z_k}{1-p_k}\Big)\prod_{k'\in\mathcal{T}'}z_{k'}\Big] = \mathbb{I}\Big(\begin{matrix}\mathcal{S}\subseteq\mathcal{S}',\\ \mathcal{T}\subseteq\mathcal{T}'\end{matrix}\Big)\prod_{j'\in\mathcal{S}'\setminus\mathcal{S}}p_{j'}\prod_{k'\in\mathcal{T}'\setminus\mathcal{T}}p_{k'}. \tag{9}$$

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

$$\mathbb{E}\Big[\prod_{j\in\mathcal{S}}\Big(\frac{z_j}{p_j} - \frac{1-z_j}{1-p_j}\Big)\prod_{k\in\mathcal{T}}\Big(\frac{z_k}{p_k} - \frac{1-z_k}{1-p_k}\Big)\prod_{j'\in\mathcal{S}'\cup\mathcal{T}'}z_{j'}\Big]. \tag{10}$$

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

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

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

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

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

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

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

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

Cases 5 and 6 ensure that (10) is non-zero only when $\mathcal{S} \subseteq (\mathcal{T} \cup \mathcal{S}' \cup \mathcal{T}')$ and $\mathcal{T} \subseteq (\mathcal{S} \cup \mathcal{S}' \cup \mathcal{T}')$, which are both necessary conditions for (9) to be non-zero. In addition, note that each j from case 7 contributing a factor of p_j to (10) also contributes at least one factor of p_j to (9). The remaining j from other cases contribute a factor of at least 1. Notably, both (10) and (9) are non-negative, with (10) dominating (9), so also the bounding the covariance. (10) will be largest when $|\mathcal{S} \cap \mathcal{T}| = \beta$ and are disjoint from $\mathcal{S}' \cup \mathcal{T}'$ such that all individuals in $\mathcal{S} \cap \mathcal{T}$ fall into case 2, allowing us to bound the covariance by $(p(1-p))^{-\beta}$. □

We are ready to bound the variance. By the law of total variance,

$$\text{Var}[\widehat{TTE}] = \text{Var}\left[\mathbb{E}\left[\widehat{TTE} \;\middle|\; \mathbf{z}\right]\right] + \mathbb{E}\left[\text{Var}\left[\widehat{TTE} \;\middle|\; \mathbf{z}\right]\right]$$

$$= \text{Var}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}) \sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i \\ |\mathcal{S}|\leq\beta}} g(\mathcal{S}) \prod_{j\in\mathcal{S}} \left(\frac{z_j}{p_j} - \frac{1-z_j}{1-p_j}\right)\right] \tag{11}$$

$$+ \mathbb{E}\left[\frac{\sigma^2}{n^2}\sum_{i=1}^{n} \left(\sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i \\ |\mathcal{S}|\leq\beta}} g(\mathcal{S}) \prod_{j\in\mathcal{S}} \left(\frac{z_j}{p_j} - \frac{1-z_j}{1-p_j}\right)\right)^2\right]. \tag{12}$$

If $\mathcal{N}_i \cap \mathcal{N}_{i'} = \emptyset$, then $Y_i(\mathbf{z})w_i(\mathbf{z})$ and $Y_{i'}(\mathbf{z})w_{i'}(\mathbf{z})$ are functions of disjoint sets of independent variables. Thus, $\text{Cov}\big[Y_i(\mathbf{z})w_i(\mathbf{z}), Y_{i'}(\mathbf{z})w_{i'}(\mathbf{z})\big] = 0$. We let $\mathcal{M}_i$ denote the set of individuals i' such that $\mathcal{N}_i \cap \mathcal{N}_{i'} \neq \emptyset$, i.e. all individuals i' that share an in-neighbors with individual i. Note that $|\mathcal{M}_i| \leq d_{\text{in}} d_{\text{out}}$. Applying the bilinearity of covariance and the triangle inequality, we can bound (11) by

$$(11) \leq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i'\in\mathcal{M}_i}\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_i \\ |\mathcal{S}'|\leq\beta}} |c_{i,\mathcal{S}'}| \sum_{\substack{\mathcal{T}'\subseteq\mathcal{N}_{i'} \\ |\mathcal{T}'|\leq\beta}} |c_{i',\mathcal{T}'}| \sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i \\ |\mathcal{S}|\leq\beta}} |g(\mathcal{S})| \sum_{\substack{\mathcal{T}\subseteq\mathcal{N}_{i'} \\ |\mathcal{T}|\leq\beta}} |g(\mathcal{T})|$$
$$\left|\text{Cov}\Big[\prod_{j\in\mathcal{S}}\left(\frac{z_j}{p_j} - \frac{1-z_j}{1-p_j}\right)\prod_{j'\in\mathcal{S}'} z_{j'} \,,\, \prod_{k\in\mathcal{T}}\left(\frac{z_k}{p_k} - \frac{1-z_k}{1-p_k}\right)\prod_{k'\in\mathcal{T}'} z_{k'}\Big]\right|.$$

Plugging in our bounds from Lemmas 2 and 3 and the definition of $Y_{\max}$, we can simplify this bound to

$$\leq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i'\in\mathcal{M}_i} {Y_{\max}}^2 \sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i \\ |\mathcal{S}|\leq\beta}} \sum_{\substack{\mathcal{T}\subseteq\mathcal{N}_{i'} \\ |\mathcal{T}|\leq\beta}} (p(1-p))^{-\beta} \quad \leq \frac{d_{\text{in}}\, d_{\text{out}}\, {Y_{\max}}^2}{n} \cdot \left(\frac{d_{\text{in}}^2}{p(1-p)}\right)^{\beta}.$$

Next we bound (12) by

$$(12) \leq \frac{\sigma^2}{n^2}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i \\ |\mathcal{S}|\leq\beta}}\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_i \\ |\mathcal{S}'|\leq\beta}} g(\mathcal{S})g(\mathcal{S}')\mathbb{E}\left[\prod_{j\in\mathcal{S}}\left(\frac{z_j}{p_j} - \frac{1-z_j}{1-p_j}\right)\prod_{j\in\mathcal{S}'}\left(\frac{z_j}{p_j} - \frac{1-z_j}{1-p_j}\right)\right]$$

$$= \frac{\sigma^2}{n^2}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i \\ |\mathcal{S}|\leq\beta}}\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_i \\ |\mathcal{S}'|\leq\beta}} g(\mathcal{S})g(\mathcal{S}')\mathbb{I}(\mathcal{S}=\mathcal{S}')\mathbb{E}\left[\prod_{j\in\mathcal{S}}\left(\frac{z_j}{p_j^2} - \frac{1-z_j}{(1-p_j)^2}\right)\right]$$

$$= \frac{\sigma^2}{n^2}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i \\ |\mathcal{S}|\leq\beta}} g(\mathcal{S})^2 \prod_{j\in\mathcal{S}} \frac{1}{p_j(1-p_j)}$$

Plugging in our bounds from Lemmas 2, the constraint that $p_i \in [p, 1-p]$, and the fact that the number of subsets of $\mathcal{N}_i$ with size at most β is bounded above by d_{in}^{β}, it follows that

$$\leq \frac{\sigma^2}{n}\left(\frac{d_{\text{in}}}{p(1-p)}\right)^{\beta}.$$