TODO

STAT 548 Qualifying Paper

Kenny Chiu

January 6, 2022

Abstract. TODO

1 Introduction

2 Notation

Throughout this report, we closely follow the notation used by Forastiere et al. (2021). Let $G = (\mathcal{N}, \mathcal{E})$ be an undirected network where $\mathcal{N}$ is a set of N units (nodes) and $\mathcal{E}$ is a set of edges (i, j). A partition $(i, \mathcal{N}_i, \mathcal{N}_{-i})$ of $\mathcal{N}$ describes unit i's neighbourhood $\mathcal{N}_i$ (the set of N_i units connected to unit i) and the set $\mathcal{N}_{-i}$ of all other units that are not i and are not in $\mathcal{N}_i$. Let $Z_i \in \{0, 1\}$ be the treatment assigned to unit i and $Y_i \in \mathcal{Y}$ the observed outcome of unit i. Denote the treatment and outcome vector for the population $\mathcal{N}$ as $\mathbf{Z}$ and $\mathbf{Y}$, respectively, and the corresponding vectors for partition $(i, \mathcal{N}_i, \mathcal{N}_{-i})$ as $(Z_i, \mathbf{Z}_{\mathcal{N}_i}, \mathbf{Z}_{\mathcal{N}_{-i}})$ and $(Y_i, \mathbf{Y}_{\mathcal{N}_i}, \mathbf{Y}_{\mathcal{N}_{-i}})$. Let $G_i = g_i(\mathbf{Z}_{\mathcal{N}_i}) \in \mathcal{G}_i$ be some known and well-specified summary g_i of the treatments in unit i's neighbourhood, and denote the vector of neighbourhood treatments for the population as $\mathbf{G}$. Depending on the size of a unit's neighbourhood, the space $\mathcal{G}_i$ may differ between units. Let $V_g = \{i \in \mathcal{N} : g \in \mathcal{G}_i\}$ denote the subset containing v_g units that have g as a possible value for the neighbourhood treatment. Let $\mathbf{X}_i \in \mathcal{X}$ be a vector of covariates for unit i that partitions into individual-level characteristics $\mathbf{X}_i^{\text{ind}} \in \mathcal{X}^{\text{ind}}$ and neighbourhood-level/aggregated individidual-level characteristics $\mathbf{X}_i^{\text{neigh}} \in \mathcal{X}^{\text{neigh}}$.

3 Proposed methodology in context of the literature

TODOIn this section, we explain the method proposed by Forastiere et al. (2021) and discuss its relevance in the context of the related literature. We also discuss its advantages over other methods used in similar contexts.

3.1 Setting, objective and method

Forastiere et al. (2021) examine the problem of performing causal inference of treatment effects from observational network data under interference. The setting is challenging for causal inference because

1. the assignment mechanism of treatments is unknown with observational data and so estimated effects may be non-causal in the presence of unmeasured confounders, and
2. naive inference methods that ignore interference may produce biased estimates.

In the randomized study literature, these issues can generally be dealt with by designing the study in such a way that the influence of confounders is minimized and that inference methods that account for interference can be used (e.g., Saveski et al., 2017; Doudchenko et al., 2020; Jagadeesan et al., 2020; Imai et al., 2021). In the observational setting, these considerations need to be addressed by the inference method in the analysis phase. Forastiere et al. consider a setting involving a binary treatment (e.g., intervention and control) and where the interference on a unit is limited to that of its immediate neighbouring units. They propose a method to estimate the causal treatment and spillover (interference) effects that yield unbiased estimates under certain assumptions.

Under the potential outcome framework, the general procedure for estimating effects is to match units into sets based on covariate values, compute the effect contrast within each matched set, and estimate the effect by the (weighted) average of the contrasts across sets. However, matching may be difficult when the space of possible covariate values is large or if there are many covariates. Forastiere et al. (2021) address this problem in their proposed method by instead matching on a defined joint propensity score $\psi(z; g; x)$ that factorizes into a neighbourhood propensity score $\lambda(g; z; \mathbf{x}^g)$ (probability of being exposed to neighbourhood treatment g given individual treatment z and relevant covariates $\mathbf{x}^g$) and an individual propensity score $\phi(z; \mathbf{x}^z)$ (probability of being assigned treatment z given relevant covariates $\mathbf{x}^z$). Note that $\mathbf{X}^g$ and $\mathbf{X}^z$ do not necessarily correspond to $\mathbf{X}^{\text{neigh}}$ and $\mathbf{X}^{\text{ind}}$, respectively, and as they may not be disjoint. The steps for their propensity-based method are as follows:

1. **Subclassify units**.
   (a) Fit a logistic regression model on the individual treatments Z_i given covariates $\mathbf{X}_i^z$, and use the model to predict the individual propensity scores $\phi(1; \mathbf{X}_i^z)$.
   (b) Partition the units into J subclasses B_j, $j \in \{1, \ldots, J\}$, based on similar estimated individual propensity scores $\hat{\phi}(1; \mathbf{X}_i^z)$ and such that each subclass is approximately balanced in the number of treated and untreated units.
2. **Estimate potential outcomes**. Let $B_j^g = V_g \cap B_j$. For each subclass B_j:
   (a) Fit some regression model on the neighbourhood treatments G_i given the individual treatments Z_i and covariates $\mathbf{X}_i^g$, and use the model to predict the neighbourhood propensity scores $\lambda(g; z; \mathbf{X}_i^g)$.
   (b) Fit some regression model on the potential outcomes $Y_i(z, g)$ given the estimated neighbourhood propensity scores $\hat{\lambda}(g; z; \mathbf{X}_i^g)$.
   (c) Estimate the dose-response function by averaging over the estimated potential outcomes for a particular level of the joint treatment, i.e.,
$$\hat{\mu}_j(z, g; V_g) = \frac{\sum_{i \in B_j^g} \hat{Y}_i(z, g)}{\left|B_j^g\right|} .$$
3. **Estimate the average dose-response function** (ADRF) $\mu(z, g; V_g) = \mathbb{E}\left[Y_i(z, g) | i \in V_g\right]$ for a particular level of the joint treatment by taking the weighted average of the estimated dose-response functions over the subclasses, i.e.,
$$\hat{\mu}(z, g; V_g) = \sum_{j=1}^{J} \hat{\mu}_j(z, g; V_g) \left(\frac{\left|B_j^g\right|}{v_g} \right) .$$
4. **Estimate** the treatment effects $\tau(g)$, overall treatment effect τ, spillover effects $\delta(g; z)$, and overall spillover effects $\Delta(z)$ by
$$\hat{\tau}(g) = \hat{\mu}(1, g; V_g) - \hat{\mu}(0, g; V_g) , \qquad \hat{\tau} = \sum_{g \in \mathcal{G}} \hat{\tau}(g) \mathbb{P}(G_i = g) ,$$
$$\hat{\delta}(g; z) = \hat{\mu}(z, g; V_g) - \hat{\mu}(z, 0; V_g) , \qquad \hat{\Delta}(z) = \sum_{g \in \mathcal{G}} \hat{\delta}(g; z) \mathbb{P}(G_i = g) .$$

Forastiere et al. (2021) show that their estimators for the treatment and spillover effects are unbiased under three assumptions, the first two of which form the Stable Unit Treatment on Neighbourhood Value Assumption (SUTNVA, a generalization of SUTVA that relaxes the no interference assumption to allow interference of immediate neighbours) and the third being an unconfoundedness assumption that says the treatment assignment mechanism is conditionally independent of the outcomes for the given set of covariates.

3.2 Comparison to the literature

The literature that examines the similar problem of causal inference in observational data under general forms of interference is still relatively new. We summarize the common approaches in this literature and discuss how the work by Forastiere et al. (2021) fits in. We also briefly highlight other work in the related literature.

3.2.1 Other approaches under the same context

As noted by Forastiere et al., the majority of the works dealing with the same context involve either inverse probability-weighted (IPW) estimators (Liu et al., 2016) or targeted maximum likelihood estimators (TMLE) (van der Laan, 2014; Ogburn et al., 2017; Sofrygin & van der Laan, 2017) for the causal treatment effect. The main advantage of the method proposed by Forastiere et al. (2021) over these two approaches is the weaker assumptions that it requires.

IPW estimators are weighted averages of the outcomes where the weights are defined with respect to a hypothetical allocation strategy (an assumed distribution over the neighbouring treatments) and a known or correctly modelled generalized propensity score. The Bernoulli allocation strategy (Tchetgen & VanderWeele, 2012) is commonly used, which assumes that each unit in the neighbourhood is treated independently with probability α and that a unit's assignment is independent of its neighbours' assignment. This assumption rules out homophily—the tendency for units with similar characteristics to form ties—which is generally a strong and unrealistic assumption to make (Shalizi & Thomas, 2011). In contrast, the estimators proposed by Forastiere et al. (2021) only use the observed neighbourhood treatments, and therefore no assumptions that explicitly rule out homophily are made (though the issue may still manifest as an unmeasured confounder if the unconfoundedness assumption does not hold for the given set of covariates). Both the IPW estimators and the estimators proposed by Forastiere et al. (2021) rely on being able to correctly model the joint propensity score.

TMLEs are obtained by maximizing the likelihood of the outcomes defined on a structural equation model. Similar to the IPW estimators, TMLE typically involves a randomization assumption (van der Laan, 2014) on the model where the conditional joint distribution of the treatment assignments factorizes into independent conditionals given the covariates for all units, and similarly for the conditional joint distribution of the outcomes given the covariates and the treatment assignments. In comparison, the unconfoundedness assumption in the method proposed by Forastiere et al. (2021) makes a weaker assumption where the outcome and treatment assignment of each unit is conditionally independent given only the covariates of that unit. The significance of these assumptions again circle back to the argument of disregarding homophily and/or other venues of confounding. It is notable that extensions of TMLE that allow for limited forms of homophily were later introduced (Ogburn et al., 2017).

More recently, approaches that can be described as extensions of Forastiere et al. (2021)'s method have been proposed. Jackson et al. (2020) proposed estimators based on propensity score matching but which also explicitly account for homophily by modeling neighbourhood treatment assignments as an incomplete information game. Sánchez-Becerra (2021) questioned the justification of the unconfoundedness assumption with respect to the constructed propensity score (which may be high-dimensional and challenging to estimate accurately), and proposed a two-step method where a defined network propensity score is first estimated and then used as inverse weights to match units.

3.2.2 Other contexts

We briefly highlight other works in the literature that consider a slightly different setting of the causal inference problem in observational data under interference. The difference in settings makes it difficult to directly compare the proposed approaches to that of Forastiere et al. (2021).

A large body of the literature examine the inference problem under the assumption of partial interference where units are partitioned into groups with no spillover effects between groups. The focus on partial interference settings seems to be primarily due to momentum of earlier works (e.g., Sobel, 2006; Hudgens & Halloran, 2008) that looked at causal inference in randomized studies with interference, in which group-randomization tends to be more practical. Examples of recent work that assume partial interference include the work by Liu et al. (2019), Barkley et al. (2020), and Qu et al. (2021). IPW estimators are commonly used in the partial interference setting as they were originally introduced for grouped observational data (Tchetgen & VanderWeele, 2012).

A small number of works consider more specific and niche contexts. For example, Toulis et al. (2018) explore the problem of treatment entanglement (where treatment assignments are assumed to satisfy certain restrictions) and proposes a propensity score-based estimator. Zigler and Papadogeorgou (2021) focus on the problem of bipartite causal inference with interference (where the treatment is applied to one unit and the outcome is measured on another) and proposes a IPW estimator.

4 Potential bias of naive estimator when unconfoundedness holds

In this section, we describe a simple example that illustrates the relevance of Theorem 2.A, Corollary 2 and Corollary 3 in the paper by Forastiere et al. (2021). Specifically, we show how for a simple network under certain assumptions, an unbiased naive estimator for the treatment effect that assumes SUTVA may be biased depending on the exact treatment assignment mechanism.

Consider some undirected network $G = (\mathcal{N}, \mathcal{E})$ where every unit is paired (has an edge) with exactly one other unit. For simplicity, we index a unit by $i \in \mathbb{N}$ for the pair and $j \in \{1, 2\}$ for the unit within a pair. Denote $Z_{ij} \in \{0, 1\}$ as the treatment assignment of unit j in pair i. Let the neighbourhood treatment G_{ij} be whether a unit's paired counterpart is treated. Therefore, $\mathcal{G}_{ij} = \{0, 1\}$ for all units in the network and $V_g = \mathcal{N}$ for all $g \in \{0, 1\}$. For convenience of notation, we drop the dependence on V_g where applicable. Let $\mathbf{X}_{ij} = X_{ij} \in \{0, 1\}$ be some covariate available for each unit, and assume that the covariates of the units in the network are generated independently with $\mathbb{P}(X_{ij} = 1) = \mathbb{P}(X_{ij} = 0) = \frac{1}{2}$. Other details, such as the space of outcomes $\mathcal{Y}$ and the size of $\mathcal{N}$, are assumed but left unspecified due to being irrelevant for the discussion. Figure 1 shows an example network.

We examine the network under two settings. In the first setting, suppose that the treatment assignment mechanism follows

$$\mathbb{P}(Z_{ij} = 1 | X_{ij}) = \frac{1}{4} + \frac{1}{2} X_{ij} \ ,$$

i.e., the probability of being treated is greater when a unit has a 1-value for the covariate. Because X_{ij} are generated independently, it follows that Z_{ij} and G_{ij} are conditionally independent given X_{ij} in this setting (they are independent unconditionally regardless of X_{ij}). An example study corresponding to this setting may be one where the paired units correspond to a pair of friends from potentially different socioeconomic backgrounds X_{ij}, and it is of interest to determine whether being comfortable discussing financial matters (Z_{ij}) has an effect on some measure of their own financial management Y_{ij}.

In the second setting, suppose that the treatment assignment follows

$$\mathbb{P}(Z_{ij} = 1 | X_{i1}, X_{i2}) = \frac{3}{4} \mathbb{1} \left[X_{i1} = X_{i2} \right] + \frac{1}{4} \mathbb{1} \left[X_{i1} \neq X_{i2} \right]$$

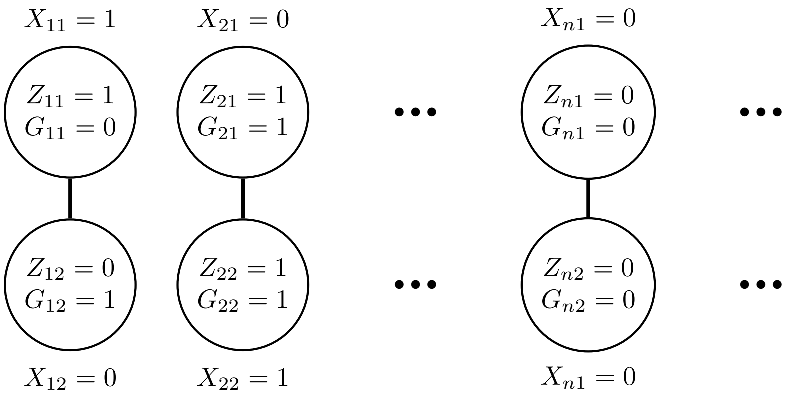

Figure 1: Example network of paired units with their individual treatment Z_{ij}, neighbourhood treatment G_{ij}, and covariate X_{ij}.

where $\mathbb{1}[\bullet]$ is the indicator function. This assignment corresponds to a type of homophily situation where the units in a pair are more likely to be treated if they share the same value of the covariate. Thus, Z_{ij} and G_{ij} are not conditionally independent given X_{ij} in this setting as the probability of a unit being treated also depends on its counterpart. An example study corresponding to this setting may be one similar to the first but involving spouses rather than friends, where spouses coming from similar backgrounds may find it easier to discuss finances (at a more intimate level).

In both settings, we further assume that Assumption 1 (no multiple versions of treatment), Assumption 2 (neighbourhood interference), and Assumption 3 (unconfoundedness) of Forastiere et al. (2021) hold. Hence, the assumptions of Corollary 1 are satisfied in the first setting, and the assumptions of Corollary 2 are satisfied in the second setting.

4.1 Setting 1: conditionally independent individual and neighbourhood treatments

In our first setting, the individual treatment Z_{ij} and neighbourhood treatment G_{ij} are conditionally independent given X_{ij}. By Corollary 1, an effect estimator that is unbiased under SUTVA will still be unbiased under the assumptions of our setting.

The overall treatment effect τ in this setting is given by

$$\begin{aligned}
\tau &= \sum_{g\in\{0,1\}} \left(\mathbb{E}\left[Y_{ij}(1,g)\right] - \mathbb{E}\left[Y_{ij}(0,g)\right]\right)\mathbb{P}(G_{ij}=g) \\
&= \frac{1}{2}\sum_{g,x\in\{0,1\}} \left(\mathbb{E}\left[Y_{ij}|Z_{ij}=1, G_{ij}=g, X_{ij}=x\right] - \mathbb{E}\left[Y_{ij}|Z_{ij}=0, G_{ij}=g, X_{ij}=x\right]\right)\mathbb{P}(X_{ij}=x) \\
&= \frac{1}{4}\sum_{g,x\in\{0,1\}} \left(\mathbb{E}\left[Y_{ij}|Z_{ij}=1, G_{ij}=g, X_{ij}=x\right] - \mathbb{E}\left[Y_{ij}|Z_{ij}=0, G_{ij}=g, X_{ij}=x\right]\right)
\end{aligned}$$

where the second equality follows from Theorem 1 in the work by Forastiere et al. (2021). The naive estimator

τ_X^{obs} that assumes SUTVA estimates the quantity

$$\begin{aligned}
\tau_X^{\text{obs}} &= \sum_{x\in\{0,1\}} \left(\mathbb{E}\left[Y_{ij}|Z_{ij}=1, X_{ij}=x\right] - \mathbb{E}\left[Y_{ij}|Z_{ij}=0, X_{ij}=x\right]\right)\mathbb{P}(X_{ij}=x)\\
&= \frac{1}{2}\sum_{x,g\in\{0,1\}} \left(\mathbb{E}\left[Y_{ij}|Z_{ij}=1, X_{ij}=x, G_{ij}=g\right] - \mathbb{E}\left[Y_{ij}|Z_{ij}=0, X_{ij}=x, G_{ij}=g\right]\right)\mathbb{P}(G_{ij}=g)\\
&= \frac{1}{4}\sum_{x,g\in\{0,1\}} \left(\mathbb{E}\left[Y_{ij}|Z_{ij}=1, X_{ij}=x, G_{ij}=g\right] - \mathbb{E}\left[Y_{ij}|Z_{ij}=0, X_{ij}=x, G_{ij}=g\right]\right)
\end{aligned}$$

where the second equality follows from the fact that $G_{ij} \perp\!\!\!\perp X_{ij}, Z_{ij}$. It then follows from the above derivations that $\tau_X^{\text{obs}} - \tau = 0$. Therefore, an unbiased estimator of the naive effect estimator is also unbiased for the treatment effect in this setting.

4.2 Setting 2: correlated individual and neighbourhood treatments

In our second setting, the covariate X_{ij} alone is insufficient for the conditional independence of the individual treatment Z_{ij} and neighbourhood treatment G_{ij}. By Corollary 2, an effect estimator that is unbiased under SUTVA will be biased under the assumptions of our setting.

The overall treatment effect τ in this setting is the same as the one given in the first setting. The naive estimator τ_X^{obs} that assumes SUTVA estimates the quantity

$$\begin{aligned}
\tau_X^{\text{obs}} &= \sum_{x\in\{0,1\}} \left(\mathbb{E}\left[Y_{ij}|Z_{ij}=1, X_{ij}=x\right] - \mathbb{E}\left[Y_{ij}|Z_{ij}=0, X_{ij}=x\right]\right)\mathbb{P}(X_{ij}=x)\\
&= \frac{1}{2}\sum_{x,g\in\{0,1\}} \left(\mathbb{E}\left[Y_{ij}|Z_{ij}=1, X_{ij}=x, G_{ij}=g\right]\mathbb{P}(G_{ij}=g|Z_{ij}=1, X_{ij}=x)\right.\\
&\quad \left. - \mathbb{E}\left[Y_{ij}|Z_{ij}=0, X_{ij}=x, G_{ij}=g\right]\mathbb{P}(G_{ij}=g|Z_{ij}=0, X_{ij}=x)\right)\\
&= \frac{1}{2}\sum_{x,g\in\{0,1\}} \left(\mathbb{E}\left[Y_{ij}|Z_{ij}=1, X_{ij}=x, G_{ij}=g\right]\sum_{x'\in\{0,1\}}\mathbb{P}(G_{ij}=g|X_{ij}=x, X_{ik}=x')\mathbb{P}(X_{ik}=x'|Z_{ij}=1, X_{ij}=x)\right.\\
&\quad \left. - \mathbb{E}\left[Y_{ij}|Z_{ij}=0, X_{ij}=x, G_{ij}=g\right]\sum_{x'\in\{0,1\}}\mathbb{P}(G_{ij}=g|X_{ij}=x, X_{ik}=x')\mathbb{P}(X_{ik}=x'|Z_{ij}=0, X_{ij}=x)\right)\\
&= \frac{1}{2}\sum_{x,g\in\{0,1\}} \left(\mathbb{E}\left[Y_{ij}|Z_{ij}=1, X_{ij}=x, G_{ij}=g\right]\sum_{x'\in\{0,1\}}\mathbb{P}(G_{ij}=g|X_{ij}=x, X_{ik}=x')\mathbb{P}(Z_{ij}=1|X_{ij}=x, X_{ik}=x')\right.\\
&\quad \left. - \mathbb{E}\left[Y_{ij}|Z_{ij}=0, X_{ij}=x, G_{ij}=g\right]\sum_{x'\in\{0,1\}}\mathbb{P}(G_{ij}=g|X_{ij}=x, X_{ik}=x')\mathbb{P}(Z_{ij}=0|X_{ij}=x, X_{ik}=x')\right)\\
&= \frac{1}{2}\sum_{x\in\{0,1\}} \left(\left(2\left(\frac{3}{4}\right)^2 + 2\left(\frac{1}{4}\right)^2\right)\left(\mathbb{E}\left[Y_{ij}|Z_{ij}=1, X_{ij}=x, G_{ij}=1\right] - \mathbb{E}\left[Y_{ij}|Z_{ij}=0, X_{ij}=x, G_{ij}=0\right]\right)\right.\\
&\quad \left. + 4\cdot\frac{1}{4}\cdot\frac{3}{4}\left(\mathbb{E}\left[Y_{ij}|Z_{ij}=1, X_{ij}=x, G_{ij}=0\right] - \mathbb{E}\left[Y_{ij}|Z_{ij}=0, X_{ij}=x, G_{ij}=1\right]\right)\right)\\
&= \sum_{x\in\{0,1\}} \left(\frac{5}{8}\left(\mathbb{E}\left[Y_{ij}|Z_{ij}=1, X_{ij}=x, G_{ij}=1\right] - \mathbb{E}\left[Y_{ij}|Z_{ij}=0, X_{ij}=x, G_{ij}=0\right]\right)\right.\\
&\quad \left. + \frac{3}{8}\left(\mathbb{E}\left[Y_{ij}|Z_{ij}=1, X_{ij}=x, G_{ij}=0\right] - \mathbb{E}\left[Y_{ij}|Z_{ij}=0, X_{ij}=x, G_{ij}=1\right]\right)\right)
\end{aligned}$$

where the third equality follows from the fact that $Z_{ij} \perp\!\!\!\perp G_{ij} | X_{i1}, X_{i2}$ and the fourth equality from Bayes' theorem manipulation (TODOAppendix?). The bias of an unbiased estimator for τ_X^{obs} is then

$$\tau_X^{\text{obs}} - \tau = \frac{1}{8} \sum_{z,x \in \{0,1\}} \left(\mathbb{E}\left[Y_{ij} | Z_{ij} = z, G_{ij} = 1, X_{ij} = x\right] - \mathbb{E}\left[Y_{ij} | Z_{ij} = z, G_{ij} = 0, X_{ij} = x\right] \right)$$

which can be verified using the bias formula in Corollary 2 given by

$$\begin{aligned}
\tau_X^{\text{obs}} - \tau = \sum_{x \in \{0,1\}} & \left(\left(\mathbb{E}\left[Y_{ij} | Z_{ij} = 1, G_{ij} = 1, X_{ij} = x\right] - \mathbb{E}\left[Y_{ij} | Z_{ij} = 1, G_{ij} = 0, X_{ij} = x\right] \right) \right. \\
& \left(\mathbb{P}(G_{ij} = 1 | Z_{ij} = 1, X_{ij} = x) - \mathbb{P}(G_{ij} = 1 | X_{ij} = x) \right) \\
& - \left(\mathbb{E}\left[Y_{ij} | Z_{ij} = 0, G_{ij} = 1, X_{ij} = x\right] - \mathbb{E}\left[Y_{ij} | Z_{ij} = 0, G_{ij} = 0, X_{ij} = x\right] \right) \\
& \left. \left(\mathbb{P}(G_{ij} = 1 | Z_{ij} = 0, X_{ij} = x) - \mathbb{P}(G_{ij} = 1 | X_{ij} = x) \right) \right) \mathbb{P}(X_{ij} = x) \\
= \frac{1}{2} \sum_{x,x' \in \{0,1\}} & \left(\left(\mathbb{E}\left[Y_{ij} | Z_{ij} = 1, G_{ij} = 1, X_{ij} = x\right] - \mathbb{E}\left[Y_{ij} | Z_{ij} = 1, G_{ij} = 0, X_{ij} = x\right] \right) \right. \\
& \left(\mathbb{P}(G_{ij} = 1 | X_{ij} = x, X_{ik} = x') \mathbb{P}(Z_{ij} = 1 | X_{ij} = x, X_{ik} = x') - \mathbb{P}(G_{ij} = 1 | X_{ij} = x, X_{ik} = x') \mathbb{P}(X_{ik} = x') \right) \\
& - \left(\mathbb{E}\left[Y_{ij} | Z_{ij} = 0, G_{ij} = 1, X_{ij} = x\right] - \mathbb{E}\left[Y_{ij} | Z_{ij} = 0, G_{ij} = 0, X_{ij} = x\right] \right) \\
& \left. \left(\mathbb{P}(G_{ij} = 1 | X_{ij} = x, X_{ik} = x') \mathbb{P}(Z_{ij} = 0 | X_{ij} = x, X_{ik} = x') - \mathbb{P}(G_{ij} = 1 | X_{ij} = x, X_{ik} = x') \mathbb{P}(X_{ik} = x') \right) \right) \\
= \frac{1}{2} \sum_{x \in \{0,1\}} & \left(\left(2 \left(\frac{3}{4} \right)^2 + 2 \left(\frac{1}{4} \right)^2 - 1 \right) \left(\mathbb{E}\left[Y_{ij} | Z_{ij} = 1, G_{ij} = 1, X_{ij} = x\right] - \mathbb{E}\left[Y_{ij} | Z_{ij} = 1, G_{ij} = 0, X_{ij} = x\right] \right) \right. \\
& \left. - \left(4 \cdot \frac{1}{4} \cdot \frac{3}{4} - 1 \right) \left(\mathbb{E}\left[Y_{ij} | Z_{ij} = 0, G_{ij} = 1, X_{ij} = x\right] - \mathbb{E}\left[Y_{ij} | Z_{ij} = 0, G_{ij} = 0, X_{ij} = x\right] \right) \right) \\
= \frac{1}{8} \sum_{x,z \in \{0,1\}} & \left(\mathbb{E}\left[Y_{ij} | Z_{ij} = z, G_{ij} = 1, X_{ij} = x\right] - \mathbb{E}\left[Y_{ij} | Z_{ij} = z, G_{ij} = 0, X_{ij} = x\right] \right)
\end{aligned}$$

where the second equality follows from the same reasoning in the derivation of τ_X^{obs}.

5 Reproducing simulation study results

In this section, we aim to reproduce the general findings of the simulation study conducted by Forastiere et al. (2021). Note that the complete Add Health dataset that Forastiere et al. work with is publicly inaccessible, and so we instead work with the Twitch Social Network dataset (Rozemberczki et al., 2021) that is comparable in size. Twitch is an American live streaming service that focuses on video game streaming. The dataset contains several networks of streamers and their mutual friendships that were collected in May 2018. We consider a hypothetical study where we are interested in understanding the effect of streamer self-promotions and advertising (the individual treatment) on the number of subscribers (users who follow a particular streamer; the outcome). It is reasonable to assume that there is interference at play where promoting one's self would inadvertently promote the streamer's network due to the site's recommendation algorithms that suggest similar streams to a viewer.

We specifically work with the EN subnetwork of the Twitch dataset, which includes a sample of streamers who stream in English. The network includes 7126 streamers and 35,324 mutual friendship relationships. A total of 3169 binary features (e.g., games liked and played, location, streaming habits, etc.) are collected from each streamer. However, the individual features are unnamed. For the purposes of this study, we only consider feature `224` and feature `569` due to their distribution (the second and third most represented features in the dataset, respectively). We view these features as covariates representing whether a streamer plays a particular `game1` and `game2`. We note that one major difference between the Twitch dataset and the Add Health dataset is that the degree of each unit is limited to at most 10 in the Add Health dataset while there is no built-in limit in the Twitch dataset. More details about the Twitch network is provided in Appendix B.

In the following sections, we describe our efforts to translate the simulation procedure and (partial) findings of Tables 1 and 2 in the work by Forastiere et al. (2021) to our Twitch data. Under various individual and neighbourhood treatment generation scenarios, Table 1 compares the theoretical bias of estimators that adjust for different sets of covariates, while Table 2 compares the observed bias and RMSE of several estimators on simulated data. In our work, we focus specifically on Scenario 1 where the unconfoundedness assumption holds given the individual-level covariates.

5.1 Treatment and outcome generation models

To study the bias of estimators, we simulate both the treatment Z (self-promotions) and the outcome Y (number of subscribers) based on the individual covariates $\mathbf{X}^{\text{ind}} = (X^{\text{game1}}, X^{\text{game2}})$ (streams game 1 and streams game 2). Our treatment generation model is given by

$$\text{logit}(P(Z_i = 1)) = \text{logit}(\phi(1; \mathbf{X}_i^{\text{ind}})) = -3 + 3X_i^{\text{game1}} + 4X_i^{\text{game2}} \, .$$

Because Z_i is generated only based on the individual covariates $\mathbf{X}_i^{\text{ind}}$, it follows that the unconfoundedness assumption holds given $\mathbf{X}_i^{\text{ind}}$. For the neighbourhood treatment G, we consider both the proportion (following Forastiere et al.) and the sum of treated neighbours (which may make more sense in our context) in our simulations. Additional details about the simulated treatments can be found in Appendix B.1.

Our outcome generation model is given by

$$\begin{aligned} &Y_i(z,g)|\mathbf{X}_i^{\text{ind}} \sim N\left(\mu(z,g,\mathbf{X}_i^{\text{ind}}), 1\right) \, , \\ &\mu(z,g,\mathbf{X}_i^{\text{ind}}) = 5 + 6\mathbb{1}[\phi(1;\mathbf{X}_i^{\text{ind}}) \geq 0.7] + 10z - 3z\mathbb{1}[\phi(1;\mathbf{X}_i^{\text{ind}}) \geq 0.7] + \beta g \end{aligned}$$

where $\beta \in \{4, 6, 8\}$ (or $\{0.4, 0.6, 0.8\}$ for sum-neighbourhood treatment) is the low, medium and high spillover effect, respectively, for proportion-neighbourhood treatment. It follows that the treatment effect $\tau(g; \mathbf{X}^{\text{ind}})$ and overall treatment effect τ are then

$$\begin{aligned} \tau(g;\mathbf{X}_i^{\text{ind}}) &= \mu(1,g,\mathbf{X}_i^{\text{ind}}) - \mu(0,g,\mathbf{X}_i^{\text{ind}}) = 10 - 3\mathbb{1}[\phi(1;\mathbf{X}_i^{\text{ind}}) \geq 0.7] \, , \\ \tau &= \sum_{x \in \mathcal{X}^{\text{ind}}} \tau(g;x)\mathbb{P}(\mathbf{X}^{\text{ind}} = x) \, . \end{aligned}$$

We remind the reader about the "super-population" perspective (Imbens & Rubin, 2015) under which expectations and probabilities are interpreted as averages over a finite network of interest. Therefore, τ is computable given a network of nodes with their covariates $\mathbf{X}^{\text{ind}}$.

5.2 Theoretical bias of estimators

We examine the bias of estimators that assume SUTVA and that adjust for differing sets of covariates. Following Forastiere et al. (2021), we consider the covariate sets $\mathbf{X}_i = \{\emptyset, \mathbf{X}_i^{\text{ind}}, \mathbf{X}_i^{z} = \mathbf{X}_i^{\text{ind}} \cup \mathbf{X}_i^{\text{neigh}}\}$ where $\mathbf{X}_i^{\text{neigh}} = \left(\frac{\sum_{k\in\mathcal{N}_i} \text{game1}_k}{N_i}, \frac{\sum_{k\in\mathcal{N}_i} \text{game2}_k}{N_i}, N_i\right)$. When no covariates are adjusted ($\mathbf{X}_i = \emptyset$), Equation (12) in Theorem 2.B. in the work by Forastiere et al. (2021) is used to compute the bias. For the other two covariate sets, Equation (11) in Corollary 2 is used. Table 1 shows the computed biases for both proportion- and sum-neighbourhood treatments on one simulated dataset.

Our findings for the estimator that does not adjust for covariates and for the estimator that adjusts for the individual covariates $\mathbf{X}_i^{\text{ind}}$ are consistent with what is reported in Table 1 in the work of Forastiere et al. (2021). For the estimator that adjusts for both individual and neighbourhood covariates $\mathbf{X}_i^{z}$, we report biases that are larger than expected. Assuming that our implementation is correct, our investigation suggests that

G exposure type	Interference (β)	Bias($\emptyset$)	Bias($\mathbf{X}_i^{\text{ind}}$)	Bias($\mathbf{X}_i^{z}$)
Proportion	Low (4)	2.937	0.057	0.541
	Medium (6)	3.009	0.085	0.812
	High (8)	3.080	0.113	1.083
Sum	Low (0.4)	3.440	-0.144	0.736
	Medium (0.6)	3.763	-0.217	1.104
	High (0.8)	4.085	-0.289	1.472

Table 1: Bias of covariate-adjusted SUTVA estimators of τ when the unconfoundedness assumption holds given $\mathbf{X}_i^{\text{ind}}$ (i.e., $Z_i \perp\!\!\!\perp G_i | \mathbf{X}_i^{\text{ind}}$) on one simulated dataset.

these large values are arising due to TODO. Because there is a much larger range in node degrees in the Twitch network compared to the Add Health network, there are many more values of the neighbourhood treatment to consider in the bias computation. This consideration combined with the smaller dataset size means that many combinations of covariate values appear in only one node TODO

5.3 Observed bias and RMSE of estimators

We examine the observed bias and RMSE of two estimators of τ that assume SUTVA on simulated datasets. The naive unadjusted estimator is the simple contrast between treated and untreated units given by

$$\tau_{\text{naive}} = \bar{Y}_{Z=1} - \bar{Y}_{Z=0}$$

where $\bar{Y}_{\bullet}$ is the mean outcome across units with $\bullet$. The regression estimator τ_{reg} (Imbens & Rubin, 2015) is extracted from a fitted linear model given by

$$\mathbb{E}[Y | Z_i, \mathbf{X}_i^{\text{ind}}] = \beta_0 + \tau_{\text{reg}} Z_i + \beta_1 X^{\text{game1}} + \beta_2 X^{\text{game2}}$$

where β_j are the other parameters of the model. The observed bias for one simulated dataset is computed as the differences between estimates $\hat{\tau}_{\text{naive}}$ and $\hat{\tau}_{\text{reg}}$ and the expected value τ, which is computed using the formula provided in Section 5.1. Table 2 reports the mean computed biases and RMSE for the two estimators over 500 simulated datasets. In each simulation, the individual and neighbourhood treatments are re-generated, and the same simulated treatments are used for all estimators and interference levels (only the outcomes are re-generated between interference levels).

		τ_{naive}		τ_{reg}	
G exposure type	Interference (β)	Bias	RMSE	Bias	RMSE
Proportion	Low (4)	2.915	2.916	0.414	0.418
	Medium (6)	2.958	2.958	0.411	0.418
	High (8)	3.004	3.005	0.411	0.422
Sum	Low (0.4)	3.553	3.556	0.448	0.504
	Medium (0.6)	3.915	3.921	0.462	0.579
	High (0.8)	4.280	4.289	0.479	0.664

Table 2: Mean bias and RMSE of estimators of τ when the unconfoundedness assumption holds given $\mathbf{X}_i^{\text{ind}}$ (i.e., $Z_i \perp\!\!\!\perp G_i | \mathbf{X}_i^{\text{ind}}$) over 500 simulated datasets.

Our findings are generally consistent with the results of Forastiere et al. (2021) reported in Table 2 (Unadjusted and Regression $\sim Z_i, \mathbf{X}_i^{\text{ind}}$ estimators under Scenario 1) where a greater interference effect leads to a greater RMSE.

TODO: DELETE

$$\begin{aligned}\delta(g;z) &= \mu(z,g,\mathbf{X}_i^{\text{ind}}) - \mu(z,0,\mathbf{X}_i^{\text{ind}}) = \beta g\\ \Delta(z) &= \delta\mathbb{E}[G_i] \qquad \forall z \in \{0,1\}\\ \mu(z,g,u) - \mu(z,0,u') &= 6(\mathbb{1}[\phi(1;u) \geq 0.7] - \mathbb{1}[\phi(1;u') \geq 0.7]) - 3z(\mathbb{1}[\phi(1;u) \geq 0.7] - \mathbb{1}[\phi(1;u') \geq 0.7]) + \beta g\end{aligned}$$

6 Extending the simulation study

TODOIn this section, we consider a small extension to the simulation study.

7 Critical appraisal and concluding remarks

TODOWe conclude this report with a critical appraisal of the method proposed by Forastiere et al. (2021).
TODOcontributions, limitations
TODOunconfoundedness assumption? Sánchez-Becerra (2021)

References

Barkley, B. G., Hudgens, M. G., Clemens, J. D., Ali, M., & Emch, M. E. (2020). Causal inference from observational studies with clustered interference, with application to a cholera vaccine study. *The Annals of Applied Statistics*, *14*(3), 1432–1448.

Doudchenko, N., Zhang, M., Drynkin, E., Airoldi, E. M., Mirrokni, V., & Pouget-Abadie, J. (2020). Causal inference with bipartite designs. *Available at SSRN 3757188*.

Forastiere, L., Airoldi, E. M., & Mealli, F. (2021). Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*, *116*(534), 901–918. https://doi.org/10.1080/01621459.2020.1768100

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

Imai, K., Jiang, Z., & Malani, A. (2021). Causal inference with interference and noncompliance in two-stage randomized experiments. *Journal of the American Statistical Association*, *116*(534), 632–644. https://doi.org/10.1080/01621459.2020.1775612

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Jackson, M. O., Lin, Z., & Yu, N. N. (2020). Adjusting for peer-influence in propensity scoring when estimating treatment effects. *Available at SSRN 3522256*.

Jagadeesan, R., Pillai, N. S., & Volfovsky, A. (2020). Designs for estimating the treatment effect in networks with interference. *The Annals of Statistics*, *48*(2), 679–712.

Liu, L., Hudgens, M. G., & Becker-Dreps, S. (2016). On inverse probability-weighted estimators in the presence of interference. *Biometrika*, *103*(4), 829–842. https://doi.org/10.1093/biomet/asw047

Liu, L., Hudgens, M. G., Saul, B., Clemens, J. D., Ali, M., & Emch, M. E. (2019). Doubly robust estimation in observational studies with partial interference. *Stat*, *8*(1), e214.

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

Qu, Z., Xiong, R., Liu, J., & Imbens, G. (2021). Efficient treatment effect estimation in observational studies under heterogeneous partial interference. *arXiv preprint arXiv:2107.12420*.

Rozemberczki, B., Allen, C., & Sarkar, R. (2021). Multi-scale attributed node embedding. *Journal of Complex Networks*, *9*(2), cnab014.

Sánchez-Becerra, A. (2021). Spillovers, homophily, and selection into treatment: The network propensity score. https://economics.sas.upenn.edu/system/files/2021-03/AlejandroSanchez_JMP_March2021_0.pdf

Saveski, M., Pouget-Abadie, J., Saint-Jacques, G., Duan, W., Ghosh, S., Xu, Y., & Airoldi, E. M. (2017). Detecting network effects: Randomizing over randomized experiments. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*.

Shalizi, C. R., & Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological methods & research*, *40*(2), 211–239.

Sobel, M. E. (2006). What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. *Journal of the American Statistical Association*, *101*(476), 1398–1407.

Sofrygin, O., & van der Laan, M. J. (2017). Semi-parametric estimation and inference for the mean outcome of the single time-point intervention in a causally connected population. *Journal of Causal Inference*, *5*(1).

Tchetgen, E. J. T., & VanderWeele, T. J. (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research*, *21*(1), 55–75.

Toulis, P., Volfovsky, A., & Airoldi, E. M. (2018). Propensity score methodology in the presence of network entanglement between treatments. *arXiv preprint arXiv:1801.07310*.

van der Laan, M. J. (2014). Causal inference for a population of causally connected units. *Journal of Causal Inference*, *2*(1), 13–74.

Zigler, C. M., & Papadogeorgou, G. (2021). Bipartite causal inference with interference. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *36*(1), 109.

A Additional derivations

Under the assumptions of our second example setting,

$$\begin{aligned}
\mathbb{P}(X_{ik} = x' | Z_{ij} = 1, X_{ij} = x) &= \frac{\mathbb{P}(Z_{ij} = 1 | X_{ij} = x, X_{ik} = x')\mathbb{P}(X_{ik} = x')}{\mathbb{P}(Z_{ij} = 1 | X_{ij} = x)} \\
&= \frac{\mathbb{P}(Z_{ij} = 1 | X_{ij} = x, X_{ik} = x')\mathbb{P}(X_{ik} = x')}{\mathbb{P}(Z_{ij} = 1 | X_{ij} = X_{ik})P(X_{ij} = x) + \mathbb{P}(Z_{ij} = 1 | X_{ij} \neq X_{ik})P(X_{ij} = x')} \\
&= \left(\frac{\frac{1}{2}}{\frac{3}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2}} \right) \mathbb{P}(Z_{ij} = 1 | X_{ij} = x, X_{ik} = x') \\
&= \mathbb{P}(Z_{ij} = 1 | X_{ij} = x, X_{ik} = x')
\end{aligned}$$

where the first equality follows from Bayes' theorem and the assumption that $X_{i1} \perp\!\!\!\perp X_{i2}$.

B Twitch dataset details

TODO
Distribution of features `224` and `569`:

	game2		
game1	0	1	Total
0	1582 (22%)	1728 (24%)	3310 (46%)
1	1834 (26%)	1982 (28%)	3816 (54%)
Total	3416 (48%)	3710 (52%)	7216

Distribution of degrees:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   2.000   5.000   9.914  11.000 720.000

 SD = 22.19026
```

B.1 Covariate balance

In one simulated dataset:

Variable	$\bar{X}_{Z=1}$	$\bar{X}_{Z=0}$	Standardized Diff.
game1	0.683	0.326	
game2	0.763	0.176	
Neighbours' game1	0.486	0.466	
Neighbours' game2	0.620	0.586	
Degree	10.757	8.717	
Proportion G_i	0.604	0.569	
Sum G_i	6.979	5.365	

Table 3: Covariate balance across individual treatment arms.

Distribution of proportion G_i's:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 0.0000  0.4444  0.6250  0.5896  0.7989  1.0000 

 SD = 0.298948
```

Distribution of sum G_i's:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  0.000   1.000   3.000   6.312   7.000 489.000 

 SD = 14.73202
```

Variable	$\bar{X}_{G\geq 0.5}$	$\bar{X}_{G<0.5}$	Standardized Diff.	$\bar{X}_{G\geq 3}$	$\bar{X}_{G<3}$	Standardized Diff.
game1	0.558	0.471		0.723	0.334	
game2	0.542	0.461		0.555	0.483	
Neighbours' game1	0.553	0.267		0.564	0.385	
Neighbours' game2	0.669	0.430		0.631	0.580	
Degree	11.694	4.897		16.905	2.432	
Z_i	0.608	0.526		0.677	0.490	

Table 4: Covariate balance across dichotomized neighbourhood treatment arms.

TODO: add histogram of log N_i and G_i?